

!!SEQUENCE LIST 1.0 ! FINDPATTERNS ON SWISS-PROT: * allowing 0 mismatches											
! <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C,T,S)(R,K,H											
SW:ANKG_PIG	CK: 8859	len: 133	finds: 1	! Q09138	sus scrofa (pig). 5'-an	SW:CHRR_RHOSH	CK: 8138	len: 213	finds: 1	! P40685	rhodobacter sphaerol
SW:ACP_BACSU	CK: 8148	len: 77	finds: 1	! P80643	bacillus subtilis. acyl	SW:COAT_TCV	CK: 1818	len: 351	finds: 1	! P06663	turnip crinkle virus
SW:ADH2_DROMO	CK: 8528	len: 253	finds: 1	! P09369	drosophila mojavensis	SW:COBT_ECOLI	CK: 7029	len: 359	finds: 1	! P36562	escherichia coli. nl
SW:ALR_TREPA	CK: 255	len: 357	finds: 1	! Q05346	treponema pallidum. ala	SW:COX3_SCHCO	CK: 7659	len: 268	finds: 1	! P14058	schizophyllum commun
SW:AMIA_SALTY	CK: 5663	len: 289	finds: 1	! P33772	salmonella typhimurium.	SW:COX4_SCHPO	CK: 7810	len: 164	finds: 1	! P79010	schizosaccharomyces
SW:AMPC_SERMA	CK: 107	len: 376	finds: 1	! P18539	serratia marcescens. be	SW:CPCL_CANPG	CK: 1605	len: 114	finds: 1	! P81580	cancer pagurus (rock
SW:AMPD_CITFR	CK: 4831	len: 187	finds: 1	! Q00831	citrobacter freundli. a	SW:CPTA_PEA	CK: 5889	len: 402	finds: 1	! P21727	pisum sativum (garde
SW:ANX2_XENLA	CK: 6452	len: 339	finds: 1	! P24801	xenopus laevis (african	SW:CSCR_ECOLI	CK: 8969	len: 331	finds: 1	! P40715	escherichia coli. su
SW:ARG1_XENLA	CK: 4321	len: 360	finds: 1	! Q91553	xenopus laevis (african	SW:CTAA_BACFI	CK: 716	len: 297	finds: 1	! Q04443	bacillus firmus. cyt
SW:ARG2_XENLA	CK: 4641	len: 360	finds: 1	! Q91554	xenopus laevis (african	SW:CTAA_BACSU	CK: 1522	len: 306	finds: 1	! P12946	bacillus subtilis. c
SW:ARG3_XENLA	CK: 4724	len: 360	finds: 1	! Q91555	xenopus laevis (african	SW:CU04_BIACR	CK: 1926	len: 127	finds: 1	! P80675	blaberus crantlifer.
SW:ARGC_ECOLI	CK: 6342	len: 334	finds: 1	! P11446	escherichia coli. n-ace	SW:CYNR_ECOLI	CK: 4304	len: 299	finds: 1	! P27111	escherichia coli. cy
SW:AT93_HUMAN	CK: 2178	len: 142	finds: 1	! P48201	homo sapiens (human). a	SW:DACX_STRSK	CK: 4817	len: 291	finds: 1	! P39042	streptomyces sp. (st
SW:ATP8_PODAN	CK: 8763	len: 50	finds: 1	! Q02653	podospora anserina. atp	SW:DBP5_YEAST	CK: 2958	len: 482	finds: 1	! P20449	saccharomyces cerevi
SW:ATP8_PONPA	CK: 3336	len: 68	finds: 1	! P92694	pongo pygmaeus abelli	SW:DEP_STRMU	CK: 9135	len: 145	finds: 1	! Q54433	streptococcus mutans
SW:ATP8_PONPY	CK: 3460	len: 68	finds: 1	! P92896	pongo pygmaeus pygmaeus	SW:DHAT_CITFR	CK: 101	len: 387	finds: 1	! P45513	citrobacter freundli
SW:ATP8_PONPY	CK: 3550	len: 68	finds: 1	! Q35584	pongo pygmaeus (orangut	SW:DHAT_KLEPN	CK: 846	len: 387	finds: 1	! Q59477	klebsiella pneumonia
SW:ATPD_HUMAN	CK: 1686	len: 168	finds: 1	! P30049	homo sapiens (human). a	SW:DHBB_HUMAN	CK: 3434	len: 310	finds: 1	! P37058	homo sapiens (human)
SW:ATP2_BACPC	CK: 4392	len: 127	finds: 1	! P09354	bacillus ps3 (thermophil	SW:DHBB_MOUSE	CK: 9317	len: 305	finds: 1	! P70385	mus musculus (mouse)
SW:BMPA_XENLA	CK: 9714	len: 398	finds: 1	! P25703	xenopus laevis (african	SW:DHBB3_RAT	CK: 3366	len: 306	finds: 1	! Q54939	rattus norvegicus (r
SW:BMPB_XENLA	CK: 1620	len: 398	finds: 1	! P30884	xenopus laevis (african	SW:DHXB_MOUSE	CK: 4788	len: 312	finds: 1	! Q70503	mus musculus (mouse)
SW:BPHB_PSEPS	CK: 1527	len: 275	finds: 1	! P08694	pseudomonas pseudocalcal	SW:DHXB_ANAPL	CK: 9157	len: 312	finds: 1	! Q57314	anas platyrhynchos
SW:BR3A_HUMAN	CK: 2610	len: 423	finds: 1	! Q01851	homo sapiens (human). h	SW:DHCB_HUMAN	CK: 391	len: 276	finds: 1	! Q75828	homo sapiens (human)
SW:BRAL_BRAFL	CK: 7365	len: 448	finds: 1	! Q17134	branchiostoma floridae	SW:DHCA_HUMAN	CK: 3202	len: 276	finds: 1	! P16152	homo sapiens carboy
SW:BRAL_BRAFL	CK: 391	len: 440	finds: 1	! P80492	branchiostoma floridae	SW:DHCA_MOUSE	CK: 204	len: 276	finds: 1	! P48758	mus musculus (mouse)
SW:BSR_BACCE	CK: 8425	len: 140	finds: 1	! P33967	bacillus cereus. blasti	SW:DHCA_RABIT	CK: 1426	len: 276	finds: 1	! P47844	oryctolagus cuniculu
SW:CADI_DICDI	CK: 3813	len: 213	finds: 1	! P54657	dictyostelium discoideu	SW:DHKA_RAT	CK: 3202	len: 276	finds: 1	! P47727	rattus norvegicus (r
SW:CAMG_MOUSE	CK: 283	len: 294	finds: 1	! P49070	mus musculus (mouse). c	SW:DHKL_STRVN	CK: 3104	len: 272	finds: 1	! P16542	streptomyces violace
SW:CDD_BACSU	CK: 9048	len: 136	finds: 1	! P19079	bacillus subtilis. cyt	SW:DHKR_STRCM	CK: 4473	len: 261	finds: 1	! P41177	streptomyces cinnamo
SW:CDD_MYCPI	CK: 6718	len: 133	finds: 1	! P47718	mycoplasma pirum. cyt	SW:DLTB_BACSU	CK: 2631	len: 252	finds: 1	! P39577	bacillus subtilis. d
SW:CDSA_HAEIN	CK: 8973	len: 288	finds: 1	! P44937	haemophilus influenzae.	SW:DSBB_VIBAL	CK: 2736	len: 165	finds: 1	! Q56578	vibrio alginolyticus
SW:CDSN_PIG	CK: 9217	len: 225	finds: 1	! Q19084	sus scrofa (pig). corne	SW:DUS1_HUMAN	CK: 5401	len: 367	finds: 1	! P28562	homo sapiens (human)
SW:CG21_SCHPO	CK: 7358	len: 415	finds: 1	! P24865	schizosaccharomyces pom	SW:DUS1_MOUSE	CK: 5871	len: 367	finds: 1	! P28563	mus musculus (mouse)
SW:CH19_DROGR	CK: 4664	len: 196	finds: 1	! P13427	drosophila grimshawi (d	SW:DUS1_RAT	CK: 6443	len: 367	finds: 1	! Q64623	rattus norvegicus (r

SW:DVR1_BRARE	ck: 2951	len: 355	finds: 1	! P35621 brachydanio rerio (zebr)
SW:EL128_DROME	ck: 68	len: 255	finds: 1	! P08761 drosophila melanogaster
SW:ELAC_TRIYU	ck: 5524	len: 102	finds: 1	! Q29143 trichostatus vulpecula (c
SW:ERP2_YEAST	ck: 5386	len: 215	finds: 1	! P39704 saccharomyces cerevisia
SW:ES1_BRARE	ck: 1627	len: 270	finds: 1	! Q90257 brachydanio rerio (zebr)
SW:EFPA_MYCLE	ck: 3764	len: 318	finds: 1	! Q35096 mycobacterium leprae. e
SW:EXBD_NEICO	ck: 4894	len: 144	finds: 1	! Q06434 neisseria gonorrhoeae.
SW:EXBD_NEIME	ck: 5831	len: 144	finds: 1	! P95376 neisseria meningitidis.
SW:FABG_ECOLI	ck: 3492	len: 244	finds: 1	! P25716 escherichia coli. 3-oxo
SW:FABG_HAEIN	ck: 4159	len: 242	finds: 1	! P43713 haemophilus influenzae.
SW:FABG_PSEAE	ck: 9922	len: 247	finds: 1	! Q54438 pseudomonas aeruginosa.
SW:FABG_VIBHA	ck: 5984	len: 244	finds: 1	! P55336 vibrio harveyi. 3-oxoac
SW:FENR_ECOLI	ck: 3071	len: 247	finds: 1	! P28861 escherichia coli. ferre
SW:FENR_SHIFL	ck: 1251	len: 135	finds: 1	! P28901 shigella flexneri. feri
SW:FIBR_AGKCO	ck: 7566	len: 203	finds: 1	! P28891 agkistrodon contortrix
SW:FIMH_SALTY	ck: 6074	len: 335	finds: 1	! P37925 salmonella typhimurium.
SW:FIH_BACSV	ck: 1150	len: 208	finds: 1	! P25449 bacillus subtilis. prob
SW:POS_AVIRK	ck: 9324	len: 322	finds: 1	! P23050 avian retrovirus nk24.
SW:POS_CHICK	ck: 7007	len: 367	finds: 1	! P11939 gallus gallus (chicken)
SW:POS_HUMAN	ck: 4677	len: 380	finds: 1	! P01100 homo sapiens (human). F
SW:FRRC_MAIRE	ck: 4642	len: 152	finds: 1	! P41347 zea mays (maize). ferre
SW:GF88_BOVIN	ck: 9403	len: 375	finds: 1	! Q18636 bos taurus (bovine). gr
SW:GF88_HUMAN	ck: 1814	len: 375	finds: 1	! Q14793 homo sapiens (human). g
SW:GF88_MOUSE	ck: 2293	len: 376	finds: 1	! Q08689 mus musculus (mouse). g
SW:GF88_PAPHA	ck: 1463	len: 375	finds: 1	! Q18828 papio hamadryas (hamadr
SW:GF88_PIG	ck: 1805	len: 375	finds: 1	! Q18831 sus scrofa (pig). growt
SW:GF88_RAT	ck: 1496	len: 376	finds: 1	! Q35312 rattus norvegicus (rat)
SW:GF88_SHEEP	ck: 1548	len: 375	finds: 1	! Q18830 ovis aries (sheep). grc
SW:GDPB_PANTR	ck: 149	len: 123	finds: 1	! Q28914 pan troglodytes (chimp)
SW:GDDA_ECOLI	ck: 4443	len: 339	finds: 1	! P37606 escherichia coli. glyce
SW:GPRD_HUMAN	ck: 2863	len: 355	finds: 1	! P49238 homo sapiens (human). f
SW:GPRD_RAT	ck: 4478	len: 354	finds: 1	! P35411 rattus norvegicus (rat)
SW:GRAB_MOUSE	ck: 92	len: 247	finds: 1	! P04187 mus musculus (mouse). g
SW:HDHA_CLOSO	ck: 1047	len: 267	finds: 1	! P50200 clostridium sordeili. f
SW:HG2A_HUMAN	ck: 9949	len: 296	finds: 1	! P04233 homo sapiens (human). h
SW:HTS8_SALTY	ck: 516	len: 359	finds: 1	! P10369 salmonella typhimurium.
SW:HTSO_ECOLI	ck: 7651	len: 228	finds: 1	! P52094 escherichia coli. histid
SW:HMPH_HUMAN	ck: 3785	len: 270	finds: 1	! Q03014 homo sapiens (human)
SW:HMPH_MOUSE	ck: 8616	len: 271	finds: 1	! P43120 mus musculus (mouse)
SW:HPRK_BACSV	ck: 4393	len: 309	finds: 1	! Q34483 bacillus subtilis. h
SW:HPRK_RHILV	ck: 1548	len: 370	finds: 1	! P28153 rhizobium leguminosa
SW:HYCD_ECOLI	ck: 675	len: 307	finds: 1	! P16430 escherichia coli. fo
SW:IAA1_WHEAT	ck: 3320	len: 124	finds: 1	! P01085 triticum aestivum (w
SW:IAA2_WHEAT	ck: 4825	len: 123	finds: 1	! P01083 triticum aestivum (w
SW:IAA5_WHEAT	ck: 4073	len: 124	finds: 1	! P01084 triticum aestivum (w
SW:ICE2_RAT	ck: 8827	len: 312	finds: 1	! P55215 rattus norvegicus (r
SW:ICW3_PSTORE	ck: 7697	len: 207	finds: 1	! P10822 psophocarpus tetrago
SW:IDI_MYCTU	ck: 9024	len: 203	finds: 1	! P72002 mycobacterium tuberc
SW:IDNR_ECOLI	ck: 2843	len: 332	finds: 1	! P39343 escherichia coli. 1-
SW:IMAN_ECOLI	ck: 6528	len: 131	finds: 1	! P09182 escherichia coli. co
SW:IMPB_SALTY	ck: 6434	len: 424	finds: 1	! P18642 salmonella typhimuri
SW:IPRA_SAGSA	ck: 1378	len: 181	finds: 1	! P31608 sagittaria sagittifo
SW:IS12_STRAL	ck: 4079	len: 256	finds: 1	! P23393 streptomyces albus g
SW:ISRA_SYNY3	ck: 4153	len: 342	finds: 1	! Q55274 synechocystis sp. (s
SW:KDD_BACSV	ck: 3484	len: 254	finds: 1	! P50842 bacillus subtilis. 2
SW:KRL_HSVS	ck: 5295	len: 467	finds: 1	! Q08097 bovine herpesvirus t
SW:LSRA_MYCTU	ck: 2319	len: 202	finds: 1	! Q10764 mycobacterium tuberc
SW:MAUF_PARBE	ck: 7474	len: 277	finds: 1	! P28897 paracoccus denitrifi
SW:MDH_RHILV	ck: 1419	len: 320	finds: 1	! Q33525 rhizobium leguminosa
SW:MESJ_HAEIN	ck: 1410	len: 430	finds: 1	! P44689 haemophilus influenza
SW:MLP6_MYXXA	ck: 8809	len: 236	finds: 1	! P38371 myxococcus xanthus.
SW:MLD6_MYCTU	ck: 4418	len: 397	finds: 1	! Q10773 mycobacterium tuberc
SW:MTDA_METERX	ck: 9524	len: 288	finds: 1	! P55818 methylobacterium ext
SW:NEUB_XENIA	ck: 1414	len: 120	finds: 1	! P43443 xenopus laevis (afri
SW:NIFY_KLEPN	ck: 8799	len: 229	finds: 1	! P09135 klebsiella pneumonia
SW:NOR1_ASPPA	ck: 7169	len: 271	finds: 1	! Q00278 aspergillus parasiti
SW:NIJM_COTJA	ck: 6471	len: 72	finds: 1	! P24968 coturnix coturnix ja.
SW:NDSM_ANCEE	ck: 1655	len: 214	finds: 1	! Q31651 anser caerulescens (
SW:NDSM_APIIL	ck: 6545	len: 167	finds: 1	! P34857 apis mellifera ligus
SW:NDSM_NEUCR	ck: 3961	len: 263	finds: 1	! P40915 neurospora crassa. n
SW:NDSM_BRAOL	ck: 9647	len: 215	finds: 1	! P42027 brassica oleracea (c
SW:NDOJ_ECOLI	ck: 2909	len: 184	finds: 1	! P33605 escherichia coli. na
SW:NDSM_NEUCR	ck: 7750	len: 201	finds: 1	! P19968 neurospora crassa. n

SW:OPPB_ECOLI	ck: 5147	len: 306	finds: 1	P31132	escherichia coli. oligo
SW:OPPB_SALTY	ck: 4920	len: 306	finds: 1	P08005	salmonella typhimurium.
SW:OPS1_CALVI	ck: 9791	len: 371	finds: 1	P22269	calliphora vicina (blue
SW:OPS1_DROME	ck: 4358	len: 373	finds: 1	P06002	drosoophila melanogaster
SW:OPS1_DROPS	ck: 7241	len: 374	finds: 1	P28678	drosoophila pseudoobscur
SW:OTCA_VIBS2	ck: 7498	len: 301	finds: 1	P96172	vibrio sp. (strain 2693
SW:OTC_SCHPO	ck: 9240	len: 327	finds: 1	P31317	schizosaccharomyces pom
SW:OXND_BOVIN	ck: 9943	len: 341	finds: 1	P31228	bos taurus (bovine). d-
SW:OXND_HUMAN	ck: 7167	len: 341	finds: 1	Q09489	homo sapiens (human). d-
SW:P2Y6_HUMAN	ck: 7545	len: 328	finds: 1	Q15077	homo sapiens (human). F
SW:PETP_RHOCA	ck: 1875	len: 166	finds: 1	P31078	rhodobacter capsulatus
SW:PEX3_YEAST	ck: 166	len: 441	finds: 1	P28795	saccharomyces cerevisiae
SW:PHBB_ALCEU	ck: 2829	len: 246	finds: 1	P14697	alcaligenes eutrophus.
SW:PHNK_ECOLI	ck: 804	len: 252	finds: 1	P16678	escherichia coli. phosp
SW:P17A_ORYSA	ck: 1305	len: 263	finds: 1	Q40708	oryza sativa (rice). ph
SW:PLAS_SYNY3	ck: 6628	len: 126	finds: 1	P21697	synechocystis sp. (str
SW:PLEK_HUMAN	ck: 1536	len: 350	finds: 1	P08567	homo sapiens (human). F
SW:PLYA_MYCPO	ck: 8743	len: 299	finds: 1	Q12639	mycosphaerella pinodes
SW:PME_ASPAC	ck: 2447	len: 331	finds: 1	Q12535	aspergillus aculeatus.
SW:PME_ASPTU	ck: 609	len: 331	finds: 1	P17872	aspergillus tubingensis
SW:PSPB_BOVIN	ck: 7861	len: 79	finds: 1	P15781	bos taurus (bovine). pu
SW:PTB_BACSV	ck: 158	len: 299	finds: 1	P54530	bacillus subtilis. prob
SW:PTI6_HUMAN	ck: 300	len: 376	finds: 1	P35237	homo sapiens (human). F
SW:PTI8_HUMAN	ck: 4145	len: 374	finds: 1	P50452	homo sapiens (human). C
SW:PURT_METTH	ck: 5550	len: 248	finds: 1	Q26272	methanobacterium thermo
SW:PYRB_ECOLI	ck: 3945	len: 310	finds: 1	P00479	escherichia coli. aspar
SW:PYRB_MYCTU	ck: 1547	len: 319	finds: 1	P71808	mycobacterium tuberculo
SW:PYRB_PYRAB	ck: 9304	len: 308	finds: 1	P77918	pyrococcus abyssi. aspa
SW:PYRB_PYRHO	ck: 9980	len: 308	finds: 1	O58451	pyrococcus horikoshii.
SW:PYRB_SALTY	ck: 2583	len: 310	finds: 1	P08420	salmonella typhimurium.
SW:PYRB_SERMA	ck: 4826	len: 305	finds: 1	P19910	serattia marcescens. as
SW:PYRB_VIBS2	ck: 5096	len: 310	finds: 1	P96174	vibrio sp. (strain 2693
SW:PYRB_BACSV	ck: 2658	len: 256	finds: 1	P25983	bacillus subtilis. dihy
SW:RAS1_DROME	ck: 2533	len: 336	finds: 1	Q27297	drosoophila melanogaster
SW:RBGR_CHRYI	ck: 4295	len: 302	finds: 1	P25544	chromatium vinosum. rut
SW:RBSD_SOLTU	ck: 4699	len: 181	finds: 1	P10647	solanum tuberosum (pot
SW:RBS1_LYCES	ck: 5240	len: 181	finds: 1	P08706	lycopersicon esculentum
SW:RBS1_PETSP	ck: 2967	len: 180	finds: 1	P04714	petunia sp. (petunia
SW:RBS1_SOLTU	ck: 4863	len: 181	finds: 1	P26574	solanum tuberosum (p
SW:RBS2_LYCES	ck: 4535	len: 180	finds: 1	P07179	lycopersicon esculen
SW:RBS3_SOLTU	ck: 4433	len: 181	finds: 1	P32764	solanum tuberosum (p
SW:RBSK_SCHPO	ck: 3214	len: 318	finds: 1	Q60116	schizosaccharomyces
SW:RBS_MUSAC	ck: 2518	len: 180	finds: 1	Q24045	musa acuminata (bana
SW:REG5_DROME	ck: 7903	len: 298	finds: 1	Q94913	drosoophila melanogas
SW:REHY_TORRU	ck: 4850	len: 218	finds: 1	P52574	tortula ruralis (mos
SW:REP2_YEAST	ck: 6596	len: 296	finds: 1	P03872	saccharomyces cerevi
SW:REPB_NEIGO	ck: 5086	len: 346	finds: 1	P37761	neisseria gonorrhoea
SW:RHL_MACMU	ck: 4371	len: 416	finds: 1	Q28849	macaca mulatta (rhes
SW:RH07_HUMAN	ck: 2901	len: 227	finds: 1	P52198	homo sapiens (human)
SW:RI87_ARCTU	ck: 4460	len: 219	finds: 1	Q28272	archaeoglobus fulgid
SW:RIPA_LUFXY	ck: 429	len: 277	finds: 1	Q00465	luffa cylindrica (sm
SW:RIPB_LUFXY	ck: 7681	len: 250	finds: 1	P22851	luffa cylindrica (sm
SW:RR21_SPIOL	ck: 718	len: 256	finds: 1	P24613	splnacia oleracea (s
SW:RR8_MARPO	ck: 6002	len: 132	finds: 1	P06362	marichanta polymorph
SW:RS13_SCHPO	ck: 3858	len: 150	finds: 1	P28189	schizosaccharomyces
SW:RS3_ACHAX	ck: 7310	len: 257	finds: 1	P41117	acholoplasma axanthu
SW:RS8_SYNY3	ck: 8926	len: 133	finds: 1	P73307	synechocystis sp. (s
SW:RVB_THBTH	ck: 3575	len: 324	finds: 1	Q56314	thermus aquaticus (s
SW:S27A_ECOLI	ck: 54	len: 217	finds: 1	P26428	escherichia coli. sl
SW:SECV_RICPR	ck: 3620	len: 433	finds: 1	Q9205	ricicetia prowazeki
SW:SIXA_ECOLI	ck: 9360	len: 161	finds: 1	P76502	escherichia coli. ph
SW:SIXA_HAEIN	ck: 2974	len: 164	finds: 1	P44164	haemophilus influenza
SW:SPAL_SALTY	ck: 7477	len: 432	finds: 1	P39444	salmonella typhimuri
SW:SPIR_SFIME	ck: 9527	len: 241	finds: 1	P21625	spiroplasma mellifer
SW:SRP_CHLPS	ck: 9648	len: 134	finds: 1	P28164	chlamydia pestifaci.
SW:SSAV_SALTY	ck: 3263	len: 352	finds: 1	P96069	salmonella typhimuri
SW:SSRI_HUMAN	ck: 8652	len: 391	finds: 1	P30872	homo sapiens (human)
SW:SSRI_MOUSE	ck: 8110	len: 391	finds: 1	P30873	mus musculus (mouse)
SW:SSRI_RAT	ck: 9240	len: 391	finds: 1	P28646	rattus norvegicus (r
SW:SSRB_CANFA	ck: 8392	len: 183	finds: 1	P23438	canis familiaris (do
SW:SSRB_HUMAN	ck: 8093	len: 183	finds: 1	P43308	homo sapiens (human)
SW:STI4_SOLTU	ck: 5285	len: 214	finds: 1	Q41495	solanum tuberosum (p
SW:SUPL_HAEIN	ck: 4	len: 311	finds: 1	P44847	haemophilus influenza

SW:SUH3_RAT	CK: 800	len: 285	finds: 1	! P50235 rattus norvegicus (rat)
SW:TRC1_CHV13	CK: 7233	len: 278	finds: 1	! P52283 chlorella virus il-3a (c
SW:TRC1_HUMAN	CK: 5007	len: 269	finds: 1	! P36402 homo sapiens (human). t
SW:TELT_HUMAN	CK: 3375	len: 167	finds: 1	! O15273 homo sapiens (human). t
SW:TELT_MOUSE	CK: 5613	len: 167	finds: 1	! O70548 mus musculus (mouse). t
SW:TRAM_AGR6	CK: 4597	len: 102	finds: 1	! Q57471 agrobacterium tumefacie
SW:TRND_SYNY3	CK: 6210	len: 231	finds: 1	! P72828 synechocystis sp. (stra
SW:TRND_SCHPO	CK: 7997	len: 415	finds: 1	! O13947 schizosaccharomyces pom
SW:TRND_YEAST	CK: 8415	len: 417	finds: 1	! O12093 saccharomyces cerevisia
SW:TRNH_DATST	CK: 5293	len: 268	finds: 1	! P50165 datura stramonium (jims
SW:TRPF_CRYNE	CK: 2342	len: 312	finds: 1	! P27710 cryptococcus neoformans
SW:TRFL_CANFA	CK: 4988	len: 371	finds: 1	! P43698 canis familiaris (dog). t
SW:TRFL_HUMAN	CK: 5439	len: 371	finds: 1	! P43699 homo sapiens (human). t
SW:TRFL_MOUSE	CK: 7622	len: 372	finds: 1	! P50220 mus musculus (mouse). t
SW:TRFL_RAT	CK: 6608	len: 372	finds: 1	! P23441 rattus norvegicus (rat). t
SW:TRF8_DENAN	CK: 8703	len: 81	finds: 1	! P01404 dendroaspis angusticeps
SW:UBC7_DROME	CK: 1891	len: 153	finds: 1	! P52487 drosophila melanogaster
SW:UMD2_ECOLI	CK: 8403	len: 139	finds: 1	! P04153 escherichia coli. umud
SW:USO2_HSVK	CK: 5996	len: 303	finds: 1	! P32517 equine herpesvirus type
SW:VANS_ENTFC	CK: 3601	len: 384	finds: 1	! Q08240 enterococcus faecium (s
SW:VE2_HPV03	CK: 4819	len: 383	finds: 1	! P36778 human papillomavirus ty
SW:VE2_HPV10	CK: 273	len: 376	finds: 1	! P36781 human papillomavirus ty
SW:VE2_HPV18	CK: 5006	len: 365	finds: 1	! P06790 human papillomavirus ty
SW:VE2_HPV28	CK: 1816	len: 376	finds: 1	! P50771 human papillomavirus ty
SW:VE2_HPV2A	CK: 734	len: 391	finds: 1	! P25482 human papillomavirus ty
SW:VE2_HPV31	CK: 9656	len: 372	finds: 1	! P17383 human papillomavirus ty
SW:VE2_HPV45	CK: 5828	len: 368	finds: 1	! P36794 human papillomavirus ty
SW:VE2_HPV51	CK: 1816	len: 358	finds: 1	! P26547 human papillomavirus ty
SW:VE2_HPV52	CK: 4491	len: 368	finds: 1	! P36796 human papillomavirus ty
SW:VEA_BPP22	CK: 7758	len: 317	finds: 1	! Q03544 bacteriophage p22. eaa
SW:VGG_BP64	CK: 4484	len: 177	finds: 1	! P03644 bacteriophage g4. major
SW:VGL1_HSV2H	CK: 501	len: 372	finds: 1	! P13291 herpes simplex virus (t
SW:VGLM_HSVSA	CK: 7071	len: 366	finds: 1	! Q01017 herpesvirus salmtili (st
SW:VIB8_AGR75	CK: 5994	len: 237	finds: 1	! P17798 agrobacterium tumefacie
SW:VIB8_AGR76	CK: 2646	len: 230	finds: 1	! P09781 agrobacterium tumefacie
SW:VIB8_AGR79	CK: 2826	len: 230	finds: 1	! P05357 agrobacterium tumefacie
SW:VID1_AGRRA	CK: 2761	len: 147	finds: 1	! P13461 agrobacterium rhizogene
SW:VID1_AGR75	CK: 2312	len: 147	finds: 1	! P18591 agrobacterium tumefa
SW:VID1_AGR76	CK: 4145	len: 147	finds: 1	! P06667 agrobacterium tumefa
SW:VDF1_NPROP	CK: 889	len: 374	finds: 1	! O10330 oryza pseudotsugata
SW:VMA2_BRGVA	CK: 6337	len: 186	finds: 1	! P29792 bovine respiratory s
SW:VMA2_HRSVA	CK: 9828	len: 194	finds: 1	! P04545 human respiratory sy
SW:VMA2_ORBVM	CK: 6559	len: 186	finds: 1	! Q84132 ovine respiratory sy
SW:VMEK_CVB	CK: 566	len: 106	finds: 1	! P3789 chrysanthemum virus
SW:VMEK_LVX	CK: 8138	len: 108	finds: 1	! P27331 lily virus x. hypoch
SW:VMEK_PVNR	CK: 3415	len: 109	finds: 1	! P17527 potato virus m (stra
SW:VP8_BPPH6	CK: 6252	len: 149	finds: 1	! P07579 bacteriophage phi-6.
SW:VPRR_SALDU	CK: 7061	len: 297	finds: 1	! P24417 salmonella dublin. v
SW:VPRR_SALTY	CK: 7032	len: 297	finds: 1	! P13041 salmonella typhimuri
SW:WN11_CHICK	CK: 4630	len: 354	finds: 1	! P49339 gallus gallus (chick
SW:WN11_COTJA	CK: 4030	len: 354	finds: 1	! P51891 coturnix coturnix ja
SW:WN11_HUMAN	CK: 3073	len: 354	finds: 1	! O96014 homo sapiens (human)
SW:WN11_MOUSE	CK: 1689	len: 354	finds: 1	! P48615 mus musculus (mouse)
SW:WN14_DROME	CK: 8394	len: 389	finds: 1	! P40589 drosophila melanogas
SW:Y021_NPROP	CK: 562	len: 298	finds: 1	! O10280 oryza pseudotsugata
SW:Y040_HUMAN	CK: 4815	len: 153	finds: 1	! Q15053 homo sapiens (human)
SW:Y209_MYCGE	CK: 590	len: 308	finds: 1	! P47451 mycoplasma genitaliu
SW:Y209_MYCPN	CK: 6533	len: 309	finds: 1	! P75485 mycoplasma pneumonia
SW:Y280_MYCGE	CK: 9511	len: 265	finds: 1	! P47522 mycoplasma genitaliu
SW:Y373_TREPA	CK: 4876	len: 477	finds: 1	! O83388 treponema pallidum.
SW:Y425_MYCPN	CK: 2944	len: 450	finds: 1	! P75172 mycoplasma pneumonia
SW:Y464_MYCPN	CK: 2836	len: 385	finds: 1	! P47702 mycoplasma genitaliu
SW:Y464_MYCPN	CK: 5866	len: 385	finds: 1	! P75112 mycoplasma pneumonia
SW:Y4EK_RHISN	CK: 5477	len: 248	finds: 1	! P55434 rhizobium sp. (strai
SW:Y4TG_RHISN	CK: 6923	len: 231	finds: 1	! P55661 rhizobium sp. (strai
SW:Y503_MERJA	CK: 6767	len: 406	finds: 1	! Q57926 methanococcus jannas
SW:Y70A_MERJA	CK: 4119	len: 102	finds: 1	! P81311 methanococcus jannas
SW:Y788_BORBU	CK: 1619	len: 440	finds: 1	! O51728 borrelia burgdorferi
SW:Y79B_MERJA	CK: 2865	len: 170	finds: 1	! P81333 methanococcus jannas
SW:Y953_MYCTU	CK: 883	len: 282	finds: 1	! P71557 mycobacterium tuberc
SW:Y988_MERJA	CK: 9981	len: 329	finds: 1	! Q58395 methanococcus jannas
SW:Y932_PSEAE	CK: 6657	len: 122	finds: 1	! P21485 pseudomonas aerugino
SW:YABP_ECOLI	CK: 98	len: 216	finds: 1	! P39220 escherichia coli. hy


```

SW:YBCC_RHOCA ck: 2127 len: 192 finds: 1 | P26166 rhodobacter capsulatus
SW:YBFT_BACSU ck: 3388 len: 249 finds: 1 | O31458 bacillus subtilis. hypc
SW:YCOB_KLEPN ck: 4280 len: 373 finds: 1 | Q48454 klebsiella pneumoniae.
SW:YCL2_MENJA ck: 2920 len: 299 finds: 1 | Q58618 methanococcus jannaschii
SW:YCB8_PORPU ck: 4681 len: 291 finds: 1 | P51321 porphyra purpurea. hypc
SW:YCBQ_ECOLI ck: 3224 len: 182 finds: 1 | P75855 escherichia coli. hypot
SW:YCG2_ECOLI ck: 1943 len: 78 finds: 1 | P51983 escherichia coli. hypot
SW:YCU1_ECOLI ck: 6021 len: 262 finds: 1 | P77333 escherichia coli. hypot
SW:YCU2_ECOLI ck: 51 len: 299 finds: 1 | P25616 saccharomyces cerevisiae
SW:YCO5_YEAST ck: 1969 len: 317 finds: 1 | P87132 schizosaccharomyces pom
SW:YDM1_SCHPO ck: 6054 len: 131 finds: 1 | O14185 schizosaccharomyces pom
SW:YDSA_SCHPO ck: 8869 len: 174 finds: 1 | O14220 schizosaccharomyces pom
SW:YDTG_SCHPO ck: 3748 len: 344 finds: 1 | P40057 saccharomyces cerevisiae
SW:YER4_YEAST ck: 5798 len: 128 finds: 1 | Q10772 mycobacterium tuberculosis
SW:YF58_MYCTU ck: 497 len: 148 finds: 1 | P31653 bacteriophage sp01. hypf
SW:YG29_BPSPI ck: 7782 len: 148 finds: 1 | P53138 saccharomyces cerevisiae
SW:YGR9_YEAST ck: 6897 len: 107 finds: 1 | Q05954 streptomyces coelicolor
SW:YGLR_STRCO ck: 9954 len: 66 finds: 1 | P10941 cryphonectria parasitica
SW:YHAI_CRYPA ck: 3269 len: 319 finds: 1 | P54604 bacillus subtilis. hypc
SW:YHCT_BACSU ck: 1929 len: 302 finds: 1 | P38756 saccharomyces cerevisiae
SW:YHG3_YEAST ck: 5569 len: 429 finds: 1 | P46851 escherichia coli. hypot
SW:YHGN_ECOLI ck: 1678 len: 197 finds: 1 | Q50604 mycobacterium tuberculosis
SW:YI29_MYCTU ck: 7074 len: 164 finds: 1 | Q10862 mycobacterium tuberculosis
SW:YU96_MYCTU ck: 1836 len: 317 finds: 1 | P39411 escherichia coli. hypot
SW:YUJX_ECOLI ck: 3873 len: 173 finds: 1 | P39430 enterobacter aerogenes
SW:YUJX_ENTAE ck: 4167 len: 54 finds: 1 | P46989 saccharomyces cerevisiae
SW:YUR8_YEAST ck: 8234 len: 196 finds: 1 | P75677 escherichia coli. hypot
SW:YXFF_ECOLI ck: 8141 len: 79 finds: 1 | P38469 marchantia polymorpha
SW:YXN7_MARPO ck: 7679 len: 69 finds: 1 | P53842 saccharomyces cerevisiae
SW:YXN6_YEAST ck: 7226 len: 139 finds: 1 | P34496 caenorhabditis elegans
SW:YXN1_CAEL ck: 5825 len: 170 finds: 1 | P34542 caenorhabditis elegans
SW:YXN3_CAEL ck: 399 len: 336 finds: 1 | P50444 caenorhabditis elegans
SW:YXN6_CAEL ck: 431 len: 381 finds: 1 | P19296 thermoproteus tenax v14
SW:YORL_TTV1 ck: 9154 len: 232 finds: 1 | Q10120 caenorhabditis elegans
SW:YOTC_CAEL ck: 9212 len: 161 finds: 1 | P14802 bacillus subtilis. hypc
SW:YOXD_BACSU ck: 2171 len: 238 finds: 1 | P77396 escherichia coli. hypot
SW:YFDC_ECOLI ck: 4986 len: 285 finds: 1 |

```

```

SW:YRPG_BACSU ck: 3316 len: 224 finds: 1 | P42981 bacillus subtilis. h
SW:YPOQ_KLEPN ck: 9681 len: 271 finds: 1 | P27509 klebsiella pneumonia
SW:YQCE_ECOLI ck: 1867 len: 425 finds: 1 | P77031 escherichia coli. hy
SW:YQEK_ECOLI ck: 8893 len: 141 finds: 1 | P77136 escherichia coli. hy
SW:YR7B_ECOLI ck: 6140 len: 169 finds: 1 | P21316 escherichia coli. hy
SW:YRBC_HAEIN ck: 1216 len: 214 finds: 1 | P45028 haemophilus influenza
SW:YRKG_BACSU ck: 1724 len: 36 finds: 1 | P54434 bacillus subtilis. h
SW:YU73_MYCTU ck: 3649 len: 118 finds: 1 | P39642 bacillus subtilis. h
SW:YWFJ_BACSU ck: 3644 len: 394 finds: 1 | P39642 bacillus subtilis. h
SW:YVVP_CAEL ck: 1176 len: 274 finds: 1 | Q93834 caenorhabditis elegans
SW:ZRP4_MAIZE ck: 6645 len: 364 finds: 1 | P47917 zea mays (maize). o-

\\End of list

Databases searched:
SWISS-PROT, Release 38.1, Released on 20NOV1999, Formatted on 28DEC1999

Total finds: 337
Total length: 29,864,866
Total sequences: 82,229
CPU time: 03:12.78

1 FINDPATTERNS on Swiss-Prot: allowing 0 mismatches
1 1 <(X)(L,200)(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)X(L,I,V,M,A,F,Y,W)(C,T,S)(R,
AAKG_PIG ck: 8859 len: 133 | Q09138 sus scrofa (pig). 5'-amp-activate
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)X(L,I,V,M,A,F,Y,W
x(175)(A)x(P)(V)xx(x)(S)(K)xx(V)xxx(A)x(42)
LVFEDTSLOKSAALVOIYELERKXVVIDPESGNLTLYTKLTFTEPPREMSKSLELOI
1:
ACP_BACSU ck: 8148 len: 77 | P80643 bacillus subtilis. acyl carrier P
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)X(L,I,V,M,A,F,Y,W
x(A)x(T)(L)xx(V)(T)(K)xx(V)xxx(G)x(60)
KADTERVTKIIVDRLGVDADVKLEASFEDLGADSLDVVELMELDEDFMEISDEDAE
1:
ADH2_DROMO ck: 8528 len: 253 | P09369 drosophila melanogaster (fruit fly)
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)X(L,I,V,M,A,F,Y,W
x(143)(P)x(P)(V)xx(A)(S)(K)xx(P)(V)xx(T)x(92)
AIAKKNIIFVAGLGIGGDSREIIVKSGPNLVITLDRIEFPAELKALNKVTVTFPYPDV
1:
INPGITKTLVHKFNSMDVPRVAELTLEHPTOTILOCAONVKAIONONGALIN
KLDIGTLEAIEWTKHMDSHI
1:
ALR_TREPA ck: 255 len: 357 | O56346 treponema pallidum. alanine racem
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)X(L,I,V,M,A,F,Y,W
x(184)(A)x(S)(A)xx(L)(C)(4)(4)xx(A)xxx(M)x(157)
MSRTARVCLIPKADAYGGADVDVAGALSCGVHSPFVAVCOESQIRAGVAPILCLSTPT
1:
STPLVHANSALICHPRAHFMVPRGGLAYGVAESVPAVRVDFVDELVTQV
RAIKKIPAGVAYQGLMRAHTEHVLPIGIVADGVNRLSPLOVCIQKMPVYVAGALCDQCVLDGFLPRTVGVDRVTELG
1:
AMTA_SALTY ck: 5663 len: 289 | P33772 salmonella typhimurium. probable
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)X(L,I,V,M,A,F,Y,W

```


1 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{139}(M)x(S)(L)xx(L)(S)(H)xx(L)xxx(M)x{211})
1: MMSNSQPFHMHPTLPBHKYPSLHSSSEARICLPTLPLOSLNLASDELTLARALAAVDH
GGGGPGGGLGSSAHPHMHSLGHLSPHAAAANMPSCLPHPGLVAAAAGHGA
AAAAAAGVAAAASAAAVGAGLASICSDTDTPRELFAERFKORIKLTOADVGSALMLKIPVGSLSOSTICFESITL
BRAF1_BRAF1 ck: 7365 len: 448 1 Q17134 branchiostoma floridae (floridae) lanc
1 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{139}(M)x(S)(L)xx(L)(S)(H)xx(L)xxx(M)x{293})
1: MMSAFTMOPTASPODOFSVSHLSAVERSEISAGSEKGDPTERDLTKLEKPLMDKFNLTNMT
VGGPDNMGWSTHTTPEMOTAFKIKYKIPFAKFLDAKERSDGD
GMEDIDOPISQGLGMLPCTGICPPPHQFAPSLGIPSHGCDRSTLNNHRSAPYPPHYORSSPPTNYGHDPAASLPMPTHDN
BRAF2_BRAF1 ck: 391 len: 440 1 P80492 branchiostoma floridae (floridae) lanc
1 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{139}(M)x(S)(L)xx(L)(S)(H)xx(L)xxx(M)x{293})
1: MKOTPSVSHLSAVERSEISAGSEKGDPTERDLTKLEKPLMDKFNLTNMT
LSTHTTAETOFIANTVYONEELTALIKIRNAPFAKFLDAKERSDGD
POFSOLGGMFLPOTGICPPPHQFAPSLGIPSHGCDRSTLNNHRSAPYPPHYORSSPPTNYGHDPAASLPMPTHDN
BSR_BACCE ck: 8425 len: 140 1 P33967 bacillus cereus. blastcidin-s deam
1 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{95}(V)x(S)(P)xx(M)(C)(R)xx(L)xxx(A)x{29})
1: MKTFNISQDDELVEVAETKIKITMLYEDNNHGAIRKTEGILISAVHIEATYGRVTCALAIAG
CADI_DICDI ck: 3813 len: 213 1 P54657 dictyostelium discoideum (slime mold
1 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{79}(L)x(S)(L)xx(L)(S)(K)xx(L)xxx(A)x{118})
1: MSVDANKKFFGNGSESEIKNGELTAFENNGDMNKKFSCVAGSVNVCNMEIDTPTPG
TTGCVTVIKDETFPKMNTVYODDNTSIFENLSEK
CAMG_MOUSE ck: 283 len: 294 1 P49070 mus musculus (mouse). calcium-signal
1 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{138}(A)x(P)(A)xx(A)(S)(H)xx(L)xxx(L)x{140})
1: MEVPATADGGERPATPGLSASORALRKRLKLNMSBORINRMINGFRPGSGSEENOTKSKPO
STREEDSPRITRIGGALLAGVAFVCKLSTIFAPLLOLAVMGLKITEFKSG
KVKKTVITRAALLSGIPAEAINRMDTYSKMGVEFDLCVYEFFTFCHELDIDYGPVP
CDD_BACSU ck: 9048 len: 136 1 P19079 bacillus subtilis. cytidine deaminase
1 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{81}(P)x(S)(P)xx(A)(C)(R)xx(L)xxx(C)x{39})
1: MNRELTETALKRDMAYAPYSKFVGAALLTKDKYKRGCNINMAYSMNCNCAERTALFKAVSE
CDD_MYCPI ck: 6718 len: 133 1 P47718 mycoplasma pitum. cytidine deaminase
1 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{83}(L)x(S)(L)xx(V)(S)(K)xx(L)xxx(L)x{164})
1: MKENDIYOKNELISNAYVPSNFRVSCILLDGGWAGVNIENSASPTICERANVSMTSG
CDSA_HAEIN ck: 8973 len: 288 1 P44937 haemophilus influenzae. phosphatidat
1 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{108}(P)x(S)(A)xx(W)(S)(K)xx(L)xxx(L)x{164})
1: MKORVLSIVILAIANALCALPTEPFALALAGVAILGIMWTOFARLKOPDLFEYTTLSGFI
GKRKLAPKVSPPKSWGVIGGLITLALVLAISFTHSNLTVAGDRNTGRTGLISVAT
VAISVIGDLTESMFRKESGVKDSOLLHGGVDRDLSLAAVPEFSYFEFVL
CDSN_PTG ck: 9217 len: 225 1 O19084 sus scrofa (pig). cornedmosin (s

1 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{37}(L)x(P)(V)xx(F)(T)(K)xx(L)xxx(S)x{172})
1: GRCPRTSVKDFGSEYVAGSSDSTVPGMTSGKITYPVGIFKDNVPVAGSGAPFAGAP
GSKSSSGHPCISVSSSTLSGPPDSQPPDSAGAKFCGSGSGKIPC
CG21_SCHPO ck: 7358 len: 415 1 P24865 schizosaccharomyces pombe (fission
1 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{122}(L)x(T)(L)xx(W)(T)(K)xx(L)xxx(S)x{277})
1: MDSTGCRATHTODENOLKNDHIVKRSKIKNTGVPAPEKAVNDINOODEPILIBONNS
LAPPNTMSVQOEDITRMLVDRNDRGIVNITLAVNLIDRELSIKV
SLOKVOVLGSLALLACKIEHIFPISYINFAVVGIFTVDEITRABRYMLMLDIDISMPGMSFLRISRAHSDIDITMLAK
CH19_DROGR ck: 4664 len: 196 1 P13427 drosophila grimshawi (fruit fly)
1 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{160}(L)x(S)(P)xx(Y)(S)(K)xx(L)xxx(A)x{20})
1: MNTATLAIIFISAGLAVGSGCGSPISYIGGPIINGLRVSISGOOSGDGMAAASAGDN
APVAKYIPONNYGSQVY
1: MNTATLAIIFISAGLAVGSGCGSPISYIGGPIINGLRVSISGOOSGDGMAAASAGDN
CHRR_RHOSH ck: 8138 len: 213 1 P40685 rhodobacter sphaeroides (rhodosphe
1 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{22}(R)x(S)(L)xx(A)(T)(H)xx(L)xxx(C)x{175})
1: MTRIRHVSADLITAYAGLISBAISLVATHTSLDCRACRAGALDAVGSIMETAPALISE
TPVAERGDLDCIATDAPLRFNSFLPLVQPFRI
COAT_TCV ck: 1818 len: 351 1 P06663 turnip crinkle virus (tcv). coat
1 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{48}(A)x(P)(V)xx(V)(T)(R)xx(L)xxx(L)x{287})
1: MENDPRKFAASDGAOMALKWKKMSITLIRBOKTRAMAGIKLSVPVOPQKTRLSAPVA
SVPTGGLITVPLDSDRFAVAGISDPKIVFGKIKMAYTGOGANDGAQISLIRVE
ITVQKRTGSTDAGIGFAGKQBPRLVSKTKTAGHEDHCHLIGNSLITFEKAPVSGLENADSPSYLGEAAGS
COAT_TOBSV ck: 3595 len: 237 1 P03598 tobacco streak virus (strain wc)
1 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{74}(P)x(P)(V)xx(V)(S)(R)xx(A)xxx(L)x{147})
1: MNTLIGGDHPSNMSRANRNNRNSRCFTGIDELIDANRKNCPAHNTVNTYRORNRNARAA
RSRFDLGGKLLNOLDDBAYWOLLDERREAKRVOIAGYWIAISKPALMPBEDFL
VNO
COBT_ECOLI ck: 7029 len: 359 1 P36562 escherichia coli. nicotinate-nucl
1 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{137}(L)x(S)(A)xx(M)(S)(R)xx(A)xxx(L)x{206})
1: MNTLIDNTIPADISTMSAORHIDGLKRPVSGKLEVLAIOLAGMPGLNGIPHYGKAV
MANTTEPAALIVSTTGRDPEEVNGVGNLDPGLFSTVAAALACOMSPAIKPYLIPSHLSAEKARIALSHLGLDEPTINMEMRIGE
DVIAKGGFDLVIGVAGLGAASGCLPVLDDGFSTVAAALACOMSPAIKPYLIPSHLSAEKARIALSHLGLDEPTINMEMRIGE
COX3_SCHCO ck: 7659 len: 268 1 P14058 schizosaccharomyces pombe (fission
1 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{156}(V)x(S)(L)xx(W)(T)(K)xx(L)xxx(S)x{196})
1: MTSIKTQSSAHAIHOSHSPMLIVSESFENLAIOTVIMHGYSHSSTTDLGLATVGSILIM
IFSFQAFEXIHAGFMSDVTFVGFSTGIGHGWLGLTFLEFVFLROVYOT
TKENHIGLETSIILYMEFVLDVLEFLVYVEMGA
COX4_SCHPO ck: 7810 len: 164 1 P79010 schizosaccharomyces pombe (fission
1 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)
x{12}(M)x(S)(W)xx(V)(S)(R)xx(A)xxx(S)x{146})
1: MFMNSMLRVSRORAANVSIYSLTRGVSSIKRNNONVAKAAAEELANAREPSDLIGPGRGD

1	CPCL_CANPG	ck: 1605	len: 114	1	P81580 cancer pagurus (rock crab). cuticle
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{54}(P)x(P)(P)x(x(V)(S)(H)xx(A)xxx(S)x{44}) QVYSGIVSPDGNNOIOTHFHSHIYLKPSGIVISDKNNQLTAGQASLOAAAPPLPVSHY			
1	CPTR_PEA	ck: 5889	len: 402	1	P21727 pisum sativum (garden pea). triose P
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{85}(P)x(P)(V)xx(F)(S)(R)xx(G)xxx(G)x{301} MESVLSRATLISLPINKLRPLANASLPYKSPFSGVSDGNNLVGROLRPELCSPLKKGAS TSNVSPAFAVAFHTVALBEPFENMAASOPLISDITLMLSLAPVIVGSMAS LTLSFNWIGFISAMISFTYRSYKRAMTMDSTNIYATISIALIYICIPALIEGPPLTKGFNDIAKVGIVKEVSDLFWY			
1	CSCR_ECOLI	ck: 8969	len: 331	1	P40715 escherichia coli. sucrose operon rep
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{24}(A)x(S)(V)xx(A)(T)(R)xx(V)xxx(I)x{291} MASLKVRLAGVSMATVSRVHNSRSPATROVDALQTLNTVVDLSARKMRAGKRPSTLAY ROPLOCFWPESALATGYRROGEADAMWRAGDLEAVGFHATGDDHITULASLN AHFSGKPDQFDVLICGNBRAAFVAYOVILARVIRPODVAWGFNLVGVHLFPLPTTIQLPHDIIGREALHIIGREGVTRI			
1	CTAA_BACFI	ck: 716	len: 297	1	Q04443 bacillus firmus. cytochrome aa3 cont
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{52}(I)x(T)(L)xx(V)(S)(H)xx(V)xxx(I)x{129} MHRKIVYVITISIVLALLOGLAVTGTGSGEGGAWPCLCFEVEIPTPAETIIEVSHRYVSG YKHSSEATLACGPFILCGOIFPCIGVCGVGHVHVGCTILFILLIMTILSR YHRIYLTWATVLSLTVAVGOTISGISTVFOTNLVGLIHALLITISFALSIVMTIIRPSH			
1	CTAA_BACSV	ck: 1522	len: 306	1	P12946 bacillus subtilis. cytochrome aa3 cc
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{51}(P)x(S)(L)xx(W)(S)(H)xx(A)xxx(S)x{123} NMKALCVLTFTFWLITLIGLAVTGTGSGCGGROMPLCHGRFPPELNPASIIEMSHREASGI TGAAVYHRESSLACGNVPLCGPLNNGDPTORHEWONGHRAALLFVWIIAAVH AITSXKDKOIFMWSICLIFITLOALSGIMIVYSELATGALAHSHFICLGVLCYFILLIARRRYSROS			
1	CU04_BLACR	ck: 1926	len: 127	1	P80675 blaberus craniifer. cuticle protein
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{103}(I)x(P)(L)xx(V)(S)(H)xx(V)xxx(V)x{8} QAVLYPSIYILSKSKVRVOELPEVGAAPADTKKAEIKEREHEVITHGLPVPYSSVVOPIVS			
1	CY2_RHOTE	ck: 2620	len: 92	1	P00098 rhodocycilus tenuis (Rhodospirillum t
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{64}(P)x(P)(A)xx(V)(T)(K)xx(A)xxx(V)x{112} ADESALAKOTKGLACHNPEKAVVGPATIGWAKTAKGQRAEKLVAKYAGGCGWAKOLGAEI			
1	CYNR_ECOLI	ck: 4304	len: 299	1	P27111 escherichia coli. cyn operon transcr
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{192}(L)x(S)(A)xx(A)(T)(R)xx(I)xxx(C)x{91} MLSHINTELAHESGVRASALHVSOPILSRQLESLGVPLFSDSGRTIRLTADGEVWRC HBOVALSRHDEKVLISAEFATREPTROIDHYCEKGAHGVIVTEANSIVALELIRR TSLSTILTPRAKITOTHDGLKRAISLAPLERAVILLRKNSMOTAAKAFHMLADKCAVVGNESR			
1	DACX_STRSK	ck: 4817	len: 291	1	P39042 streptomyces sp. (strain k15). d-ala
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C x{23}(A)x(P)(A)xx(V)(T)(R)xx(I)xxx(G)x{252} MRLRRATAVITTGALLAAGLIGATPATAVATKPTLAAVGAAMNNGTGLTYTKADTRSTGSTI			
1	DBP5_YEAST	ck: 2958	len: 482	1	P20449 saccharomyces cerevisiae (baker's
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{146}(A)x(S)(L)xx(L)(T)(R)xx(V)xxx(S)x{1320} MSDTRKPADLSTIKTIDNEKTSKTPKVPKQPEKTADSIKPAKLVPRVKEKTKQED TLEVYQMGKFTTITGOLIVPDSFEKKNINNOYVGPYGLDMRKMLQLOKI KIFVLEADNMLDQGGDCCIKNFELPKDQVLFSATRADAVAKQIVRANULTLOTNEVNDIAKQLYMDCKNEADKF			
1	DEP_STRMU	ck: 9135	len: 145	1	Q54433 streptococcus mutans. dna/pantoct
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{45}(P)x(T)(L)xx(L)(S)(K)xx(V)xxx(V)x{84} MTKIIILVSGSIAAYKADLSHQLTKLGYHNVEMTNAKQFIPLTLQVLSKNPVSYNWK			
1	DHAT_CITFR	ck: 101	len: 387	1	P45513 citrobacter freundii. 1,3-propane
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{143}(A)x(T)(A)xx(V)(T)(R)xx(V)xxx(T)x{128} MSYRMEPYLVPNVNFGRPNASIVSGRCKLGGKRALVTDGGLRAIKDGAVDRTLHLREAG INDPLMLGRPAPLITATGMDALTHAVEVYISKRDANPVTDAALQALILIRNDRQ AVALGSNIKARENNAYASLLAGMAFNANNLGYHAMAHQDGLXDMPHGVANAVLLPHVARYNLIANPEKFIADIAFEGENTDGL			
1	DHAT_KLEPN	ck: 846	len: 387	1	Q59477 klebsiella pneumoniae. 1,3-propan
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{143}(A)x(T)(A)xx(V)(T)(R)xx(V)xxx(T)x{128} MSYRMEPYLVPNVNFGRPNASIVSGRCKLGGKRALVTDGGLRAIKDGAVDRTLHLREAG INDPLMLGRPAPLITATGMDALTHAVEVYISKRDANPVTDAALQALILIRNDRQ AVALGSNIKARENNAYASLLAGMAFNANNLGYHAMAHQDGLXDMPHGVANAVLLPHVARYNLIANPEKFIADIAFEGENTDGL			
1	DHB3_HUMAN	ck: 3434	len: 310	1	P37058 homo sapiens (human). estradiol 1
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{193}(L)x(S)(M)xx(A)(S)(K)xx(V)xxx(S)x{101} MGOVLEOFPLILGLLGVCLVLCVRSRCVLLVWVLEPKSLRSMGMAVITGAGGIGKRA GLINTISSGIALPMPPLSYLSASAKAFVVFESKALNVEYRDGIIIOVLTLPSTVSTPMTK AMTKYNTVYTTADEVFESLKVYTGSGTCGLAHETIATILIPAMAYSAFOFLLTHYVAVLTKNTKVR			
1	DHB3_MOUSE	ck: 9317	len: 305	1	P70385 mus musculus (mouse). estradiol 1
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{189}(L)x(S)(L)xx(A)(S)(K)xx(V)xxx(S)x{100} MEKLFIAAGFVGLVCLVLCVKMRSSRLFLFCFALGSLRSMGMAVITGAGGIGKAYSFE NISSGAALRPMPLSYLSASAKAFVVFESKALNVEYRDGIIIOVLTLPSTVSTPMTK YLNKMTKRTADEVFESLKVYTGABSCCLAHETIATILNRIPRIEYSTAQRFLTRYSDYLRKNISNR			
1	DHB3_RAT	ck: 3366	len: 306	1	O54939 rattus norvegicus (rat). estradio
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{189}(L)x(S)(L)xx(A)(S)(K)xx(V)xxx(S)x{101} MEKLFIAAGFVGLVCLVLCVKMRSSRLFLFCFALGSLRSMGMAVITGAGGIGKAYSFE NISSGCVRPMPLSYLSASAKAFVVFESKALNVEYRDGIIIOVLTLPSTVSTPMTK YLNKMTKRTADEVFESLKVYTGABSCCLAHETIATILNRIPRIEYSTAQRFLTRYSDYLRKNISNR			
1	DHBR_MOUSE	ck: 4788	len: 312	1	O70503 mus musculus (mouse). putative st
1	1:	<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W) x{161}(V)x(S)(V)xx(V)(T)(R)xx(L)xxx(V)x{133} MECAPAGGLVYVWAGSTAVLAFVDFSCLEHYKSGIFGVOSVWY ERSKGVILNTISSAGMLPVPLLTIVSATAPVDFSCLEHYKSGIFGVOSVWY LVATIKLAKIOPPLIDRPSAETFEVSKATVGFQTRTGTGVVHSLMGSSINSIMPMWYFKIIMGFSKSLRNRYLKRRKN			


```
ESRLTWLSTIIIIIIIAVISIAOVLIIQFLFTGRKONY
ESL_BRARE   ck: 1627  len: 270   | Q90257 brachydanio rerio (zebrafish) (zebra
1:          <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{56}(A)x(T)(M)xx(A)(T)(K)xx(A)xxx(L)x{198}
1:          MASRALAKAQAAMALYQPCIMHGCDMGNGNNGNINIAVFGCGCMWGCTIHEAATMHLSTN
LLACRVPSTLVTNGYERDESSRWGRPNTNRYAKVSMGARNTREPEYAYDEK
NKVISTPTFMWETDHYHIFDGIQNNVKAHMYRTAK
ETFA_MYCLE  ck: 3764  len: 318   | O33096 mycobacterium leprae, electron trans
1:          <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{171}(P)x(P)(A)xx(A)(T)(K)xx(A)xxx(A)x{131}
MAELALVHEHTEGALKRVSALITRAAVLGEPAAYVVGPGTSAPLVGDLKTAGAAKIYVAESDA
ATKTTARAPAVYDNRPDITETATVYVVGSGRGVGSADNPSEALADSLGAAGASRA
AYDSGYTFGQFOIGCTKITSPLYIALGSGAIOHRAGMGTKITVAYANKDEAPIFEIADFVGSDFKVAPOLDGKAKRG
EXBD_NEIGO  ck: 4894  len: 144   | O06434 nelsseria gonorrhoeae, biopolymer tr
1:          <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{33}(M)x(T)(M)xx(L)(T)(H)xx(P)xxx(P)x{95}
MAFGSMNSDSDSPMSDINVTPLVDVMLVLIVFMITWPLVTHSIPLELPTASEQANKODKOPKDP
1:          MAFGSMNSDSDSPMSDINVTPLVDVMLVLIVFMITWPLVTHSIPLELPTASEQANKODKOPKDP
EXBD_NEIME  ck: 5831  len: 144   | P95376 nelsseria meningitidis, biopolymer t
1:          <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{33}(M)x(T)(M)xx(L)(T)(H)xx(P)xxx(P)x{95}
MAFGSMNSDSDSPMSDINVTPLVDVMLVLIVFMITWPLVTHSIPLELPTASEQANKODKOPKDP
1:          MAFGSMNSDSDSPMSDINVTPLVDVMLVLIVFMITWPLVTHSIPLELPTASEQANKODKOPKDP
FABG_ECOLI  ck: 3492  len: 244   | P25716 escherichia coli, 3-oxoacyl-[a
1:          <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{110}(L)x(S)(V)xx(L)(S)(K)xx(M)xxx(L){118}
NMFECKIALVTGASGSGIRAIETIARAGAKVIGTATISENGAQISIDYLGANGKGLMNTVDPAST
1:          NMFECKIALVTGASGSGIRAIETIARAGAKVIGTATISENGAQISIDYLGANGKGLMNTVDPAST
VYAPGFETDTRALSDQAGILAIQVPAGRIGAGELINAAVFLASDEAYITGE
TLHNGSGMYM
FABG_HAEIN  ck: 4159  len: 242   | P43713 haemophilus influenzae, 3-oxoacyl-[a
1:          <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{108}(L)x(S)(V)xx(L)(S)(K)xx(L)xxx(T)x{118}
MOGKIALVTGSGRGIRAIETIARAGAKVIGTATISENGAQISIDYLGANGKGLMNTVDPAST
1:          MOGKIALVTGSGRGIRAIETIARAGAKVIGTATISENGAQISIDYLGANGKGLMNTVDPAST
APGFATDMEVLIDEGAGILSNVPAGRIGAKIKAVAFILASDDAGYITGTL
HYNGGLYIS
FABG_PSEAE  ck: 9922  len: 247   | O54438 pseudomonas aeruginosa, 3-oxoacyl-[a
1:          <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{113}(L)x(S)(L)xx(L)(S)(K)xx(L)xxx(T)x{118}
MSLQKRVAILVVGASGSGIQAIALIELGRIGAVIGTATISENGAQISIDYLGANGKGLMNTVDPAST
1:          MSLQKRVAILVVGASGSGIQAIALIELGRIGAVIGTATISENGAQISIDYLGANGKGLMNTVDPAST
TVNAVAPGFIDTDRRELPEAOREALIGDIPGLRIGAEIEAKVGVFLASDGAAYV
TCATVPVNGMYMS
FABG_VIBHA  ck: 5984  len: 244   | P55336 vibrio Harveyi, 3-oxoacyl-[acyl-carb
1:          <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{110}(L)x(P)(I)xx(M)(S)(K)xx(L)xxx(M)x{118}
NMLEKRIALVVGASGSGIRAIETIARAGAKVIGTATISENGAQISIDYLGANGKGLMNTVDPAST
1:          NMLEKRIALVVGASGSGIRAIETIARAGAKVIGTATISENGAQISIDYLGANGKGLMNTVDPAST
TLHNGSGMYM
FENR_ECOLI  ck: 3071  len: 247   | P28861 escherichia coli, ferredoxin--nadp r
1:          <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{15}(V)x(P)(L)xx(A)(T)(R)xx(P)xxx(L){121}
MTSTVTTIVGCGGLPVRPLSTATRGPRRCARVAAAGADASNDKSVEMRKFSSEQYARSNT
```

GDF8_BOVIN ck: 9403 len: 375 1 O18836 bos taurus (bovine). growth/different
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 1: KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 GHDLAATFPPGGEDGLNPLEVAVTIDTPKRSRDRDGLDDEHSTESRCRPLTVDFEAFGMDWIIAPKRYKANCSCGECEFFVLOK
 GDF8_HUMAN ck: 1814 len: 375 1 O14793 homo sapiens (human). growth/different
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 1: KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 GHDLAATFPPGGEDGLNPLEVAVTIDTPKRSRDRDGLDDEHSTESRCRPLTVDFEAFGMDWIIAPKRYKANCSCGECEFFVLOK
 GDF8_MOUSE ck: 2293 len: 376 1 O08689 mus musculus (mouse). growth/different
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 1: KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 GHDLAATFPPGGEDGLNPLEVAVTIDTPKRSRDRDGLDDEHSTESRCRPLTVDFEAFGMDWIIAPKRYKANCSCGECEFFVLOK
 GDF8_PAPHA ck: 1463 len: 375 1 O18828 papio hamadryas (hamadryas baboon).
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 1: KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 GHDLAATFPPGGEDGLNPLEVAVTIDTPKRSRDRDGLDDEHSTESRCRPLTVDFEAFGMDWIIAPKRYKANCSCGECEFFVLOK
 GDF8_PIG ck: 1805 len: 375 1 O18831 sus scrofa (pig). growth/different
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 1: KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 GHDLAATFPPGGEDGLNPLEVAVTIDTPKRSRDRDGLDDEHSTESRCRPLTVDFEAFGMDWIIAPKRYKANCSCGECEFFVLOK
 GDF8_RAT ck: 1496 len: 376 1 O35312 rattus norvegicus (rat). growth/different
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 1: KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 GHDLAATFPPGGEDGLNPLEVAVTIDTPKRSRDRDGLDDEHSTESRCRPLTVDFEAFGMDWIIAPKRYKANCSCGECEFFVLOK
 GDF8_SHEEP ck: 1548 len: 375 1 O18830 ovis aries (sheep). growth/different
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 1: KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 GHDLAATFPPGGEDGLNPLEVAVTIDTPKRSRDRDGLDDEHSTESRCRPLTVDFEAFGMDWIIAPKRYKANCSCGECEFFVLOK
 GLPB_PANTR ck: 149 len: 123 1 Q28914 pan troglodytes (chimpanzee). glycosylase
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 1: KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 GHDLAATFPPGGEDGLNPLEVAVTIDTPKRSRDRDGLDDEHSTESRCRPLTVDFEAFGMDWIIAPKRYKANCSCGECEFFVLOK
 GPD_A_ECOLI ck: 4443 len: 339 1 P37606 escherichia coli. glycerol-3-phosphate
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 1: KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 GHDLAATFPPGGEDGLNPLEVAVTIDTPKRSRDRDGLDDEHSTESRCRPLTVDFEAFGMDWIIAPKRYKANCSCGECEFFVLOK
 x{68}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{68}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{68}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 1: KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 GHDLAATFPPGGEDGLNPLEVAVTIDTPKRSRDRDGLDDEHSTESRCRPLTVDFEAFGMDWIIAPKRYKANCSCGECEFFVLOK
 GPRD_HUMAN ck: 2863 len: 355 1 P49238 homo sapiens (human). probable 5
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 1: KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 GHDLAATFPPGGEDGLNPLEVAVTIDTPKRSRDRDGLDDEHSTESRCRPLTVDFEAFGMDWIIAPKRYKANCSCGECEFFVLOK
 GPRD_RAT ck: 4478 len: 354 1 P35411 rattus norvegicus (rat). probable
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 1: KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 GHDLAATFPPGGEDGLNPLEVAVTIDTPKRSRDRDGLDDEHSTESRCRPLTVDFEAFGMDWIIAPKRYKANCSCGECEFFVLOK
 GRAB_MOUSE ck: 92 len: 247 1 P04187 mus musculus (mouse). granzyme b
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 1: KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 GHDLAATFPPGGEDGLNPLEVAVTIDTPKRSRDRDGLDDEHSTESRCRPLTVDFEAFGMDWIIAPKRYKANCSCGECEFFVLOK
 HDNA_CLOSO ck: 1047 len: 267 1 P50200 clostridium sorcellii. 7-alpha-hy
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 1: KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 GHDLAATFPPGGEDGLNPLEVAVTIDTPKRSRDRDGLDDEHSTESRCRPLTVDFEAFGMDWIIAPKRYKANCSCGECEFFVLOK
 HG2A_HUMAN ck: 9949 len: 296 1 P04233 homo sapiens (human). hla class 1
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 1: KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 GHDLAATFPPGGEDGLNPLEVAVTIDTPKRSRDRDGLDDEHSTESRCRPLTVDFEAFGMDWIIAPKRYKANCSCGECEFFVLOK
 HIS8_SALTY ck: 516 len: 359 1 P10369 salmonella typhimurium. histidine
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 1: KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 GHDLAATFPPGGEDGLNPLEVAVTIDTPKRSRDRDGLDDEHSTESRCRPLTVDFEAFGMDWIIAPKRYKANCSCGECEFFVLOK
 HISO_ECOLI ck: 7651 len: 228 1 P52094 escherichia coli. histidine trans
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 1: KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 GHDLAATFPPGGEDGLNPLEVAVTIDTPKRSRDRDGLDDEHSTESRCRPLTVDFEAFGMDWIIAPKRYKANCSCGECEFFVLOK
 HMPH_HUMAN ck: 3785 len: 270 1 O03014 homo sapiens (human). homeobox pr
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{65}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 1: KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 KPMKDGRTYGTGIRSLKIDMNPSTGIMOSIDVATVQNNLAKOPESNLGIEIKALDEN
 GHDLAATFPPGGEDGLNPLEVAVTIDTPKRSRDRDGLDDEHSTESRCRPLTVDFEAFGMDWIIAPKRYKANCSCGECEFFVLOK

KRL_HSVBS CK: 5295 len: 467 1 Q08097 bovine herpesvirus type 1.2 (strain
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{143}(A)X(P)(A)XX(V)(T)(R)XX(A)XXX(A)X(308)
 1: MERAERLACQACATWRSPRCCCAAEPSGSRGCGSVRAAARCAAGSADLYLVNNEGPE
 x{143}(A)X(P)(A)XX(V)(T)(R)XX(A)XXX(A)X(308)
 VEAATGAPAOEHVVLKIGASASTLAEMERLIDHDHNVKIKAVYFHELYCAVL
 ARYREDLTHLMKIDRPMALPTALOVTRAVYQGLATIHRSKRIRHROVKENFVNLNPGVCLGDGGAHGPVTEPRYGLAGLTETNS
 LSPA_MYCTU CK: 2319 len: 202 1 Q10764 mycobacterium tuberculosis. probable
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{123}(P)X(S)(P)XX(A)(T)(R)XX(V)(K)XX(A)XXX(A)X(1238)
 1: MPDEPTGSADPLTSEAGAGEPNAPAPRKLMLSLVAVVLLDITVTKVAVOQLPQGPVSE
 FGFDEPTVGRHADGDTVGRKADG
 MAUF_PARDE CK: 7474 len: 277 1 P29897 paracoccus denitrificans. methylamit
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{123}(P)X(S)(P)XX(A)(T)(R)XX(V)(K)XX(A)XXX(A)X(1238)
 1: MVSIEDHIGLSAGQASVPPCKLFPQSPSARLAVLLAALAGAGVALASAGQPLMAYVIGAA
 KNYVIGLIGLIDYLVYVOTPLVMTLAVFTGNIAIAIVALENLGRFLPV
 AVNALPDPFRVQAMLARHOENALADGAILTALGAGFTVALI
 MOH_RHILV CK: 1419 len: 320 1 O33525 rhizobium leguminosarum (biovar vici
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{196}(L)X(T)(M)XX(V)(T)(K)XX(L)XXX(I)X(1108)
 1: MARKKILIGISGMIGTLAIGLKEGLDVLFDIADGIPQCKGLDISQSSPVEGFPVNTGASD
 DTMVPLARSTVGGIPLTDLYMTGKLELREDEGAEVGLKTSAAV
 APAASALEMASTLAKRVLPCAHLSGQYKMDXVGPVIGAGVERIEIDNTEKREAFDRKSVAGVAGLCEACINIAPALK
 MESI_HAEIN CK: 1410 len: 430 1 P44689 haemophilus influenzae. putative cel
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{157}(L)X(P)(L)XX(E)(T)(R)XX(L)XXX(A)X(1257)
 1: MDLSEDKOJOKATNATLALISGLDGLVLLSPAKLCKOKPHLPPLSTIRAHIHGSLPNDS
 LNMVDESNEEDRNFRLNEILPELREHAFDPLAVOSGHSQHCFOOCLINDLL
 SEIFTEHCIOIKNOERLCOFROYSLAKOTALLRMLAENOLEMNSKROLINDVIFPAKEANPOQLVNRVIRRYODSLYLKPSFI
 MLPA_MYXXA CK: 8809 len: 236 1 P38371 myxococcus xanthus. putative lipopro
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{83}(L)X(P)(A)XX(A)(S)(R)XX(A)XXX(A)X(240)
 1: MTRKNIVNLTALVAGSLILTCNTEPOPTNCFVDESSMAVKDIDVDSKDNANGDECTTATLVEI
 YVRAVWPAPCDIATSTPAENCNGVSGINPEFAVCOPTSAITGTCVPRADIP
 SFK
 MMU6_MYCTU CK: 4418 len: 397 1 Q10773 mycobacterium tuberculosis. putative
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{141}(P)X(T)(P)XX(I)(S)(H)XX(A)XXX(A)X(240)
 1: MOGISVGLKRGWAVRVSVDITDIDGDLAAGVLAIVTDLKLAIDPOLVALLPDELISQDINE
 AATFKDIOGATYDILLIAGIALSLILIMLITRILAVIGVLAISGAGSG
 LSVIYVHLLIGLQIVYVLAVALVILLAVGSDYNLLISREKEEIGALNTGITRANAGTGVVTAAGLFAATSSSEVFDLRLVOC
 MTDA_METEX CK: 9524 len: 288 1 P55818 methylobacterium extorquens. methylis
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{174}(A)X(T)(A)XX(A)(S)(R)XX(A)XXX(A)X(98)
 1: MSKLLSOFDDATPSYDVVAVGVDGAGHITGYGNTVDNAGVADGTITRGRKROSTALFVG
 ADDSRAEAYKGAHFYVTAAGLIELLELLOAAMNNESSIEIVADYNAOPILGIGID
 ATDKREIGRRRAGLIGIKIKLRACIKLFPSSSEGVADAEIYKLAKEWA
 NEUB_XENIA CK: 1414 len: 120 1 P43443 xenopus laevis (african clawed frog)
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{106}(L)X(S)(V)XX(A)(T)(H)XX(V)XXX(G)X(79)
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{180}(P)X(T)(A)XX(V)(S)(K)XX(A)XXX(V)X(175)
 1: MNGSLQHOEKLSTFIDGPEDETVLYLNGASRGIGRGLIEAFIORPKSTVAVMLRWVRTAT
 MSNDTIFMRMLAFOSLPDIOPOIYDWMIAOESGEBTLPERLATLTPQCLAASFPSATAVMS
 1: MSAVPLTRMLPLRFLTHLLLSFPLVFCMEFSESDANIEKIRGNQMAIGHFMKREKLOD
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{8}(M)X(P)(L)XX(L)(T)(H)XX(L)XXX(I)X(96)
 1: MSVAVPLTRMLPLRFLTHLLLSFPLVFCMEFSESDANIEKIRGNQMAIGHFMKREKLOD
 NIFY_KLEPN CK: 8799 len: 229 1 P09135 klebsiella pneumoniae. nify prote
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{60}(V)X(S)(P)XX(W)(S)(R)XX(A)XXX(G)X(153)
 1: MSNDTIFMRMLAFOSLPDIOPOIYDWMIAOESGEBTLPERLATLTPQCLAASFPSATAVMS
 1: MSAVPLTRMLPLRFLTHLLLSFPLVFCMEFSESDANIEKIRGNQMAIGHFMKREKLOD
 NORA_ASPPA CK: 7169 len: 271 1 Q00278 aspergillus parasiticus. aflatoxi
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{180}(P)X(T)(A)XX(V)(S)(K)XX(A)XXX(V)X(175)
 1: MNGSLQHOEKLSTFIDGPEDETVLYLNGASRGIGRGLIEAFIORPKSTVAVMLRWVRTAT
 MSNDTIFMRMLAFOSLPDIOPOIYDWMIAOESGEBTLPERLATLTPQCLAASFPSATAVMS
 1: MSAVPLTRMLPLRFLTHLLLSFPLVFCMEFSESDANIEKIRGNQMAIGHFMKREKLOD
 NUTM_COTUA CK: 6471 len: 72 1 P24968 coturnix coturnix japonica (japan
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{8}(L)X(S)(I)XX(A)(T)(K)XX(L)XXX(S)X(148)
 1: SFLSPPELSITLTKVLLSSFLMIRASTPRFOTDMLMLKKNLPLTLMCLMHTS
 NUTM_ANCE CK: 1655 len: 214 1 Q31651 anser caerulescens (goose). nadh-
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{188}(P)X(T)(M)XX(I)(T)(K)XX(A)XXX(G)X(110)
 1: GSVLQCSLNTSYLNTMALTLMTAFATAYSTIRMTLTVAGOTRILPPWMSKNNENPLITAP
 ELOVAAKSAATLSMTGKIKAYLSFALSITVILMTH
 NUTM_APLI CK: 6545 len: 167 1 P34857 apis mellifera ligustica (common
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{157}(L)X(P)(L)XX(E)(T)(R)XX(L)XXX(A)X(1257)
 1: MDLSEDKOJOKATNATLALISGLDGLVLLSPAKLCKOKPHLPPLSTIRAHIHGSLPNDS
 LNMVDESNEEDRNFRLNEILPELREHAFDPLAVOSGHSQHCFOOCLINDLL
 SEIFTEHCIOIKNOERLCOFROYSLAKOTALLRMLAENOLEMNSKROLINDVIFPAKEANPOQLVNRVIRRYODSLYLKPSFI
 NUTM_NEUCR CK: 3961 len: 263 1 P40915 neurospora crassa. nadh-ubiquinol
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{23}(P)X(T)(I)XX(I)(S)(R)XX(L)XXX(A)X(224)
 1: MATKLEPFLMRIRARARTRKSTSTRIAPVSRSDTLWVRNTPDNNDIPKFE
 LFTFIEVEGLGACANAPVAVOINDYFEDLPETIKOVLAKESVIVDSKAPQEP
 OSGROTCEANAAGLITSLTSEPYGPVTRSDL
 NUTM_BRAOL CK: 9647 len: 215 1 P42027 brassica oleracea (cauliflower).
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{36}(A)X(T)(P)XX(I)(T)(H)XX(P)XXX(X)X(163)
 1: MAINTTNTTRPLVLOSHRAANVSHLITSIPALSPATITSTSTRGPPSTSAAPPGLSTAE
 VDIYVGCPTFAELLYGLLOLQKKINRKNDELFAMNK
 NUCO_ECOLI CK: 2909 len: 184 1 P33605 escherichia coli. nadh dehydrogen
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{13}(P)X(T)(L)XX(I)(T)(H)XX(P)XXX(X)X(163)
 1: MEFAFICGLIATLRLVITHTNPHALVLLISLALISGVSFSLGAYRAGALEIYVAGMI
 NUTM_NEUCR CK: 7750 len: 201 1 P19968 neurospora crassa. nadh-ubiquinol
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{106}(L)X(S)(V)XX(A)(T)(H)XX(V)XXX(G)X(79)
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{180}(P)X(T)(A)XX(V)(S)(K)XX(A)XXX(V)X(175)
 1: MNGSLQHOEKLSTFIDGPEDETVLYLNGASRGIGRGLIEAFIORPKSTVAVMLRWVRTAT
 MSNDTIFMRMLAFOSLPDIOPOIYDWMIAOESGEBTLPERLATLTPQCLAASFPSATAVMS
 1: MSAVPLTRMLPLRFLTHLLLSFPLVFCMEFSESDANIEKIRGNQMAIGHFMKREKLOD
 <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
 x{8}(M)X(P)(L)XX(L)(T)(H)XX(L)XXX(I)X(96)
 1: MSVAVPLTRMLPLRFLTHLLLSFPLVFCMEFSESDANIEKIRGNQMAIGHFMKREKLOD

Mon Feb 14 08:07:43 2000

sp.cat

Page 14

```
1: MASKVTGVVKTAGGVVSVOKTTVOSVGVWIRIRRAIDPNRNSGVLPVYRNKSPGSLDPI
MDVYKQLAEPSISEAYPCRSFS
OPBP_ECOLI ck: 5147 len: 306 i P31132 escherichia coli. oligopeptide trans
1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{185}(L)x(S)(I)x(X)(I)(T)(R)xx(W)xxx(L)x{105}
MKFTLRCLDAIPDLTLLITISFMRRLAPGSGFTGERALPEVLANTEAKYHLNDPIMTQFSY
1:
MVALSLAYIASIRITRGSMTIEVLSHFIFITAKKLPRIIRFIRALPALLPV
SYMGPATVGIITGSMVLETTIGLPJIGOLFVNGALNRDYSVLVLTITLVGALTITLFINAIVDLVAVIDPKIRY
OPBP_SALTY ck: 4920 len: 306 i P08005 salmonella typhimurium. oligopeptide
1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{185}(L)x(S)(I)x(X)(I)(T)(R)xx(W)xxx(L)x{105}
MKFTLRCLDAIPDLTLLITISFMRRLAPGSGFTGERALPEVLANTEAKYHLNDPIMTQFSY
1:
MVALSLAYIASIRITRGSMTIEVLSHFIFITAKKLPRIIRFIRALPALLPV
SYMGPATVGIITGSMVLETTIGLPJIGOLFVNGALNRDYSVLVLTITLVGALTITLFINAIVDLVAVIDPKIRY
OPSI_CALVI ck: 9791 len: 371 i P22269 calliphora vicina (blue blowfly) (ca
1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{179}(L)x(P)(V)xx(W)(S)(R)xx(P)xxx(L)x{176}
MERSTPLIGSPFALINGSVTDKVTDPMAHLVHPYMNOPFAPMPKAFLAAYMLITISMGCN
1:
MTLAPVPGMSRYVEGNLTSCGIDYLEERDMNPRSYLFIYSIFYVYLPFLICYW
FIIAVSAHEKAMREQAKKNVKSLSRSESDASGEGKLAKVALVTISLFWAMPYITINTLIGFEGLTPIINTMGACFAKSAO
OPSI_DROME ck: 4358 len: 373 i P06002 drosophila melanogaster (fruit fly).
1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{181}(L)x(P)(V)xx(W)(S)(R)xx(P)xxx(L)x{176}
MESPAVAAOQIGQFAPALINGSVTDKVTDPMAHLVHPYMNOPFAPMPKAFLAAYMLITISMGCN
1:
SIMCLAPRGMRSRYVEGNLTSCGIDYLEERDMNPRSYLFIYSIFYVYLPFLICYW
YMTITIAVSAHEKAMREQAKKNVKSLSRSESDASGEGKLAKVALVTISLFWAMPYITINTLIGFEGLTPIINTMGACFAKSAO
OPSI_DROPS ck: 7241 len: 374 i P28678 drosophila pseudoobscura (fruit fly)
1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{182}(L)x(P)(V)xx(W)(S)(R)xx(P)xxx(L)x{176}
MDSFAAVATQIGQFAPALINGSVTDKVTDPMAHLVHPYMNOPFAPMPKAFLAAYMLITISMGCN
1:
TIMCCCLAPVPGMSRYVEGNLTSCGIDYLEERDMNPRSYLFIYSIFYVYLPFLICYW
SYMFIITIAVSAHEKAMREQAKKNVKSLSRSESDASGEGKLAKVALVTISLFWAMPYITINTLIGFEGLTPIINTMGACFAKSAO
OTCA_VIBS2 ck: 7498 len: 301 i P96172 vibrio sp. (strain 2693). ornithine
1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{3}(L)x(S)(V)xx(L)(T)(H)xx(A)xxx(G)x{232}
MENLISYDLSKQOITLILALRVRANPAPYQALAGSIVITIIYKPSLRTRVFDIGIRKLGS
1:
CPRESSDAIIVQAAALAIISGKINVDNDLAIYVITIGDINVSQDDPPLA
OYKXKIPFIOINALKRTIKIVLHCOFAHRELEITSEVMOGESHLDIOENRMAHONAVLLTLK
OTC_SCHPO ck: 9240 len: 327 i P31317 schizosaccharomyces pombe (fission y
1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{9}(L)x(S)(I)x(X)(I)(T)(R)xx(W)xxx(L)x{302}
MSFKFPRHLISIDLSRGEIVKLIDBSSEIPIKONFONRONSVOGSLSSONAMFISKRSTRT
1:
MIANAAGVITHTSVAKPDDVAVRDIDISIVAEANENGSTFEIIVNDPKAVKNADIY
VTDIWMIGOEAEKREOLKQFTGFOYTGIEIMKLAKSCFKHCLPRHPEVSDVEFYGENSLVQEAENRKTIVAVLEALVNRGEI
OXDD_BOVIN ck: 9943 len: 341 i P31228 bos taurus (bovine). d-aspartate ox
1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{15}(L)x(T)(A)xx(L)(S)(R)xx(P)xxx(L)x{310}
MDVRIAVAGAGVGLSTACISKMPGCSITITISDFTFETISDVAAAGMLIPPTPTIOKQKQ
1:
IYNVCSGLSROJLAGSGKIFPVRGVYKLVQAFVWYKHFIDSSGLTITIFGVSNTL
1: GSTRQKGMNLSPEAEISKEILLSRCALPELSRGAYDLREKVGRLPRTPSVYLEKELLADDSRLPVVHHYGHSGGSLMHGTA
OXDD_HUMAN ck: 7167 len: 341 i Q99489 homo sapiens (human). d-aspartate
1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{15}(L)x(T)(A)xx(L)(S)(R)xx(P)xxx(L)x{310}
MDVRIAVAGAGVGLSTACISKMPGCSITITISDFTFETISDVAAAGMLIPPTPTIOKQKQ
1:
IYNVCSGLSROJLAGSGKIFPVRGVYKLVQAFVWYKHFIDSSGLTITIFGVSNTL
GSTRQKGMNLSPEAEISKEILLSRCALPELSRGAYDLREKVGRLPRTPSVYLEKELLADDSRLPVVHHYGHSGGSLMHGTA
P2Y6_HUMAN ck: 7545 len: 328 i Q15077 homo sapiens (human). p2y purinoc
1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{119}(L)x(P)(P)xx(A)(T)(K)xx(L)xxx(A)x{133}
MEMDNGTGALGIPPTCYRRENTGQLLPYSAVLAAGLPNLCVITTOICTSRALTRAV
1:
YDLSPPALATHPTFMALTVIGFLPFRALACLCLACLRKODGPAEVRQER
RGAARAAVVAAPAFISFLPHTKTAIVRSTPGVCTVLEAFMAAKGTIRFPAANSVLDPLFTYFTOKKRRRPHELLQK
PETP_RHOCA ck: 1875 len: 166 i P31078 rhodobacter capsulatus (rhodosphe
1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{68}(L)x(S)(V)xx(V)(T)(K)xx(L)xxx(L)x{82}
MADTGAPGETLITFLITDEQLRRGIAMFEFAVAGFTADPDRLDODHGYGRAHRAIRHFINREBP
1:
PEX3_YEAST ck: 166 len: 441 i P28795 saccharomyces cerevisiae (baker's
1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{145}(L)x(S)(L)xx(L)(T)(R)xx(L)xxx(L)x{280}
MAPNORSGLDRHRRKGLISITGLAALFTGSVYFFVKKMLKQOURLTEGHFIEQIKR
1:
NCKKIONREYVWTSMWSPDEKADDAVMAAKSKKXGCOVYINPDAFISTSMWT
LNGWLSYNEIITNOLEIFEDDHRDLTLEESSRLTINFRNTNQIIOQNNNTLITLLKRDSSGQEFLLSQTLDDALTSF
PHBB_ALCEU ck: 2829 len: 246 i P14697 alcaligenes eutrophus. acetate
1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{112}(L)x(S)(L)xx(V)(T)(R)xx(L)xxx(A)x{118}
MTORIAVVGGMGIGITAIQCLADGFRVAVAGCGPNSPRREKMLEOQKALGFDFIASGNVA
1:
VNTSPGVIATDMVKAIRQDVYDKIVATIPVKRLGPEELIASICAWLSSESGFST
GADFSLNGGLHNG
PHNK_ECOLI ck: 804 len: 252 i P16678 escherichia coli. phosphonates tr
1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{4}(L)x(S)(V)xx(L)(T)(H)xx(A)xxx(G)x{232}
MNQPLISVNNLHLVAPKSGFSDFDIMPGEVIGIVGESSGKTLTILKSASALTPOQGEIH
1:
VSVQARLLDLGLGVEINLAVIYTHLQVARRLLADRLVAKRGQVYESGLIDRV
LDPPHPIYQOLVSSVLQ
P17A_ORISA ck: 1305 len: 263 i Q40708 oryza sativa (rice). p17a protei
1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{52}(V)x(S)(L)xx(Y)(S)(R)xx(L)xxx(A)x{195}
MEDGKHEFVFGVGLGAGAWCVRVVAARAGHRMALDMAAGAHVPAKRADEVLSLEYSRPL
1:
RPGNTYIDPIMKDELLELLEGVGVSVKRVFLVANDDASSDEBMIDLSPGVETL
EELAGADHAMKSKRELCDLLRITAAKYD
PLAS_SYNZ3 ck: 6628 len: 126 i P21697 synechocystis sp. (strain pcc 680
1:
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{77}(A)x(T)(A)xx(L)(S)(H)xx(L)xxx(A)x{333}
MSKFFITLGLLIVSFLSVSPAAANATYTKMOSDSGALVFESTYTITRAGEYKWN
1:
PLEK_HUMAN ck: 1536 len: 350 i P08567 homo sapiens (human). plectstin
```


ESETSYVKVCLDGPVFKAQEVAL

RA51_DROME ck: 2533 len: 336 ! Q27297 drosophila melanogaster (fruit fly).

```

1  <X>[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(
    x{45}(V,X(A)(V'xx(A)(T)(K)xx(L)xxx(P,x[275]
1:  MERTUVOAQOQEEBEEBEGP,SVYTKLIGSGTITNADIKLQOAS,SHYBESYANAKRKQLMALPGJGGG
NESEYLVNVAFTRAINSDOOTKILIMAGMGEFSTYALIDYSAALRSDITDGS
PLARONLIGLFLRLOPLAEFGVAIVITNOVYNSLDGAPFAPKPIGHHIABHSTTRLYLRKKGETRIKCIYDSPCLPESE

```

1 RBRC_CHRVI ck : 4295 len: 302 | P25544 chromatium vinosum, rubisco operon
<X>[1..200](L,I,V,M,P)X(P,T,S)(L,I,V,M,P)XX(L,I,V,M,A,F,Y,W)(C
>[83](E,X)(L,I,V,M,P)(S)(R,X)(L,I,V,M,X)[203]
1: MYSTPQIVFVFNVAHNSHTPAALFELHLSQPAVSQVRLDEBILGSLFERGKQVLTAGRE
GERA1SLARAEETVTEEGSGTQNAHRRPFSSENGQITRBGMQNTBNDAYQAVR
SGGSGVSVLTIETLEETRTVADVEQFPDRQMVLVTRGKULSPAGAFKFEVTLSEAAHMCRLG

```

RBSO_SOUTU  ck: 4699  len: 181  ! P10647 solanum tuberosum (potato). ribulose
1:
1:
<X>[1,200](t,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)((
x(S)(V)x(S)(A)xx(A)(T)(R)xx(V)xxx(S)X(L16)
MSSISAAVATRSNVAQSMAPFTGLGSLFKTKNNNVDDITLSLNGRGYRCMOVPEPI

```

```

RBS1_LYCES      ck: 5240   len: 181    | P08706 Lycopersicon esculentum (tomato). r
1              <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
                x{5}(V)x{s}(A)xx(A)(T)(R)xx(V)xxx(S)x{l60}
                MASTVSAAATRSNVAQASWAPFTGLKSAASFPTTKNNNVDTISLASNGRGRCMQWPPII
                . 1:

```

```

RBS1_PETSP ck: 2967 len: 180 i P04714 petunia sp. (petunia). ribulose bisp
1
<X>{1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)((
X(5)(I)X(S)(A)Xx(A)(T)(R)Xx(V)Xxx(S)x(15)
MASSVSSAAVAATRTNVAQASMTAPENGKLSAVSPVSSKQNDLITSIASNGGRVOCQMVPEYGA
1:

```

```

1      RBS1_SOUTU  ck : 4863  len: 181  ! P26574 solanum tuberosum (potato) . ribulose
      1:
      <X>[1,200]{L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)((
      x{5}(I)x(S)(A)xx(A)(T)(R)xx(V)xxx(G)x(160)
      MSSVSSAANVADRTNVTQAGSMIAFTGLKSAATFPYSKQKLDITSINSGRVRRCMGVMPPII

```

```

RBS2_LYCES      ck: 4535      len: 180      | P07179 lycopersicon esculentum (tomato). r
1
1:
1:  <(X){1,200}<(L,I,V,M,A,P)X(P,T,S)<(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)<(
    x{5}(I)x(S)(A)xx(A)(T)(R)xx(V)xxx(S){1159}
    MASSVTSAAVATRRSNTQSMVAPFTGLKSSITFEPTKQNDITISINSNGRVCVMQVPPINN

```

```

RBS3_SOLTU  ck: 4433   len: 181   ! P32764 solanum tuberosum (potato). ribulose
1
<X>{1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)((
x(5)(V)x(S)(A)xx(A)(T)(R)xx(V)xxx(S)x(16)
MASSIVSAAVATRSNVAAQASMTAPFTGLKSASFVYTKNNNVNDITSLASNGRVRCMQWPPIN
1:

```

```

1      RBSK_SCHPO  ck: 3214  len: 318   ! O60116 schizosaccharomyces pombe (fission y
      <(X)[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
      x{169}(P)x(P)(A)xx(L)(S)(H)xx(I)xxx(A)x{133}

```

1 MINIVLGSNPTDLMVRKPIGSGEITHGSDGDESTGNCGKANQAVAVARLSNPADTRYSM
HDMISVCALVYPNEEHAILLNOADSPTELENDVAASKRLLFCFGRKAVITITGSO
GAYIKSANGEBALVSACKKAYDITAAODTFIGAEFNSITAHGQPKDSEFAAKCSALTIVQRKAASSISLELEVDSSTNLKANT

```

RBS_MUSAC  CR: 2518  len: 180  ! 024045 musa acuminata (banana). ribulose
1          <X>[1,200]{L,I,V,M,A,P)X(P,T,S){L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
          x{6}{V)x(s){A)xx(E){(T){R)xx(P)xxx(S){158}
          MYSSMYSSAATFTTAASTAAQSSMVAFTTOLKASASAFPTYTRKNADLSHLPSNGRYQCAKATNP
1:
TGX

```

```

REGD_ROME  CR: 7903  Len: 298  | Q94913  drosoophila melanogaster (fruit fly)
1
<X>[1..200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
1
1:
MTTAAATACACCLGCGGAGTGGT(G)T(G)X(X,L)XX(X,A)X(I,I)
DKATCYGQSSSELEVAALALSEPGRKFLTLVLRAPDSPPAEEDMNPFLPDE
DLPTLSSWSGRGSAQAQODRNNLTKAALNLTVAERPPPTPAEDPSPVAGVGRSDDPOA

```

```

REHL_TORRU  CR: 4850  Len: 218  1 P52574 tortula ruralis (moss).  rehlydrh.

1
<X>(1,200)(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
x{160}(L)x(S)(L)xx(A)(S)(K)xx(L)xxx(A)x{42}
1:
MGGGAAAGDLPVDPDQADSSMGHAKNDYKDCDWTIIIFSPGDIYPPVCTTELGRIAYNFEFEKK
KQKGEPPVYISPSVSEKKAQKQFPGQWETVNLKALRMTFVD

```

```

REP_YEAST ck: 6596 len: 296 | P03872 saccharomyces cerevisiae (baker's
1
<X>(1,200)(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
16)(I,X)(C,R)X(C)I(XX)(A,X)264
MODIFIERACTKARVAVYVQVQVCFIMIAVDVIDIESRRKDELLFFVYVIRPMSESLTT
1
LVSATVQLLTLYVPRQALAGDLPFIPSLKIDINSFNEELAAENRLOOKKEFLGCR
TEVNHANTNEVPSRRISRDITNMGALONTIIEGCKAVPTFKRKAIVVRKSKSNSTRV

```

1
 <X>{1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,M
 X{160}(P)X(X(A)(S)(K)X(X(A)(XXX(X)X{170)
 1: MOTEGKNKLLVGGAGFISAYVRRHILQNTBRSVNLDELITAGNLESTLTDADNPVRYAEQV
 AMQRTPLSTVSNNSNGRPPRPFKPLMLVNLATLLEELBAGKQKIDWLF
 FEDHARALYQVTTSEVSEETINIGNHNKLTPLPPLDARPGVARGEDLITFYODRPGHARVAVDAKIRRDG

```

RHL_MACMU CR: 4371 len: 416 1 Q28845 macaca mulatta (rhesus macaque).

1
<(x){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
x{108}(L)x(S)(I)xx(A)(T)(R)xx(M)xxx(I)x{129}
1:
SSRYRCC(C)P(L)ITLAEALILFEFFYYTSLDEOKCSAYOCDDITMAVVLGLDF
GGTAAACDLPRLPGLGEGEDKXYOTLPSLFLMGLPMPFNLSLNLPIER
KAAVSTYALAVASVTLIVSSSLAHPGKRLNMTLMAHAAAGVLAASCHVSLIPAVLVLVGLTISIGAKCLPVCFFNV

```

```

1      RHO7_HUMAN  ck: 2901  len: 227   ! P52198 homo sapiens (human) . rho-related
      <X>[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,M
      x[126](L)x(T)(L)xx(L)(S)(K)xx(L)xxx(T)x[85]
      1:  MEGSGRCRHHRLQRLRDSRGKGRNSQLSGRDRGNEGSIHMDRAKSCILM
      TVASIGRGRHRLQRLRDSRGKGRNSQLSGRDRGNEGSIHMDRAKSCILM

```

```

RUB7_ARCFU  ck: 4460  len: 219  1 028272 archaeoglobus fulgidus. putative
1
<X>[1,200](L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)X(X,L,I,V,M,A,F,Y,W
X[101](A)X(T)(L)X(X)(S)(R)X(X,P)XXX(Y)X[102]
1:
MRYIVVNAASIDGKISDSKRLNISCDEEDRIYDRRAASDAIMNGIGIVLADDPRLTYK
PIFGGRDSPVACDSESLKCRKRIEYIRIGEGFAVTRFNR

```

1: MRYVAVNAASLDGKISDESKQRLRISCEEDLRIVDRIRASDAIMWIGIVLADDPRLTYK
PFIIGRDSPTVCDGESEFLKCRLEKIERIGEGFAVTAFFNR

[illegible]

1	Y47G_RHISN	ck: 6923	len: 231	1	P55661	<i>Rhodobium</i> sp. (strain ngr234).	probe
1							
1	Y503_METJA	ck: 6767	len: 406	1	O57926	<i>Methanococcus jannaschii</i> .	putative 2
1							
1	Y70A_METJA	ck: 4119	len: 102	1	P81311	<i>Methanococcus jannaschii</i> .	hypothetic
1							
1	Y788_BORBU	ck: 1619	len: 440	1	O51728	<i>Borrelia burgdorferi</i> (Lyme disease s	
1							
1	Y79B_METJA	ck: 2865	len: 170	1	P81233	<i>Methanococcus jannaschii</i> .	hypothetic
1							
1	Y933_MYCTU	ck: 883	len: 282	1	P71557	<i>Mycobacterium tuberculosis</i> .	hypothet
1							
1	Y98B_METJA	ck: 9981	len: 329	1	O58395	<i>Methanococcus jannaschii</i> .	hypothetic
1							
1	YA32_PSEAE	ck: 6657	len: 122	1	P21485	<i>Pseudomonas aeruginosa</i> .	hypothetical
1							
1	YABP_ECOLI	ck: 98	len: 216	1	P39220	<i>Escherichia coli</i> .	hypothetical 24.8
1							
1	YBCC_RHOCA	ck: 2127	len: 192	1	P26166	<i>Rhodobacter capsulatus</i> (rhodospneu	
1							
1	YBFT_BACSU	ck: 3368	len: 249	1	O31458	<i>Bacillus subtilis</i> .	hypothetical 2
1							
1	YC08_KLEPN	ck: 4280	len: 373	1	O48454	<i>Klebsiella pneumoniae</i> .	hypothetic
1							
1	YC21_METJA	ck: 2920	len: 299	1	O58618	<i>Methanococcus jannaschii</i> .	hypothet
1							
1	YC38_PORPU	ck: 4681	len: 291	1	P51321	<i>Porphyra purpurea</i> .	hypothetical 3
1							
1	YCBQ_ECOLI	ck: 3224	len: 182	1	P75855	<i>Escherichia coli</i> .	hypothetical f1
1							
1	YCG2_ECOLI	ck: 1943	len: 78	1	P75991	<i>Escherichia coli</i> .	hypothetical 8.
1							
1	YCOI_ECOLI	ck: 6021	len: 262	1	P51983	<i>Escherichia coli</i> .	hypothetical 28
1							
1	YCJZ_ECOLI	ck: 51	len: 299	1	P77333	<i>Escherichia coli</i> .	hypothetical tr
1							

[illegible]

1 YHAI_CRYPA ck: 3269 len: 319 | P10941 cryphonectria parasitica (cheesnut
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
x{162}(L)x(T)(P)xx(V)(C)(K)xx(V)xxx(L)x{1241}
MAOLRKSOISIVLESVDPTTVPFVSVKXEEVAPGCTITTWERDSCGDVPGPLSHGDLRL
1: MAOLRKSOISIVLESVDPTTVPFVSVKXEEVAPGCTITTWERDSCGDVPGPLSHGDLRL
TTGCVARADLTLLCWNGVRSGTSQIEKSAADHYHVVAEYQSEDDGLFQ
ALUGLAEDPLARIGLNFLNLAEPNAGSLAREPVTPOVTRKSTRTMTGDRPTIVRMQWYGHQHQCSCGCGVEEERFP

1 YHCT_BACSTU ck: 1929 len: 302 | P54604 bacillus subtilis. hypothetical 3
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
x{144}(L)x(T)(A)xx(V)(S)(K)xx(L)xxx(M)x{262}
NMOKRGLELILNPPGKQTRVTHFKYKSNCKRPFYSLTALKASVPIQDWMSSHOQIVYNNESVLNMNIYKGGDR
1: NMOKRGLELILNPPGKQTRVTHFKYKSNCKRPFYSLTALKASVPIQDWMSSHOQIVYNNESVLNMNIYKGGDR
RHOIRVHLASGHLITGDSLTGGSKLNNQALHANKVQAPPTIDELIVAEAPPADMKNLCTYES

1 YHG3_YEAST ck: 5569 len: 429 | P38756 saccharomyces cerevisiae (baker's
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
x{196}(L)x(S)(M)xx(A)(T)(K)xx(P)xxx(S)x{217}
MANNTKLTATLATSIVSTLIDIASVWKEVLSGCAANKNKTVSRPRYDHLFREDLANVAF
1: MANNTKLTATLATSIVSTLIDIASVWKEVLSGCAANKNKTVSRPRYDHLFREDLANVAF
NIESKYDILVYAHNKKIDIVSNGVATKSPDRSALDISEMTEDPSRCVRRKLR
KRQIATGISVVSNNELDPRDDIISPDCHERNAVNRDEALRHLPELGTMGIGLSIATWILTKVSGIPKENEYANKLKY

1 YHGN_ECOLI ck: 1678 len: 197 | P46851 escherichia coli. hypothetical 21
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
x{120}(L)x(T)(L)xx(L)(S)(H)xx(L)xxx(G)x{61}
NMELISAVLILIMDPLGNLDPFMSVLKHEPEKRRRAIRVELLIALVWLFLEAGEKILA
1: NMELISAVLILIMDPLGNLDPFMSVLKHEPEKRRRAIRVELLIALVWLFLEAGEKILA

1 YI29_MYCTU ck: 7074 len: 164 | Q50604 mycobacterium tuberculosis. hypot
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
x{49}(V)x(P)(P)xx(L)(T)(H)xx(L)xxx(L)x{199}
MGEVAVVGINVEOPONOPVILLNANGDKRLITWIGSSAAMALIEQGVBPBPRLTHDLIRD
1: MGEVAVVGINVEOPONOPVILLNANGDKRLITWIGSSAAMALIEQGVBPBPRLTHDLIRD

1 YJ96_MYCTU ck: 1836 len: 317 | Q10862 mycobacterium tuberculosis. hypot
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
x{111}(V)x(T)(L)xx(I)(S)(R)xx(A)xxx(L)x{190}
MSAOITNGIVAGDGSCHSHTAEVNAADQNRNALVVOVPEVITAPBGNAFEFSRQE
1: MSAOITNGIVAGDGSCHSHTAEVNAADQNRNALVVOVPEVITAPBGNAFEFSRQE
GSPSTGAELAPFEASRGVDYVATLHANSMDPLFPFLNAPLEMRNLDEDEK
MLARLSCWDPRBYDVVHVHVVYVCDPRAPRLLELAQTALQVNVGSHGRGCFPGMHLGSVSAVNVSGAPYIYAKIPQDAVYA

1 YJX1_ECOLI ck: 3873 len: 173 | P39411 escherichia coli. hypothetical 18
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
x{144}(L)x(T)(V)xx(V)(R)xx(L)xxx(A)x{13}
MLTHQVCATTNPAKQIALQALHAFHEIFGEGSCHIASVAVESGVPEOPGSEBETAGANRYVA
1: MLTHQVCATTNPAKQIALQALHAFHEIFGEGSCHIASVAVESGVPEOPGSEBETAGANRYVA

1 YJX1_ENTAE ck: 4167 len: 54 | P39430 enterobacter aerogenes (aerobacte
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
x{24}(V)x(T)(A)xx(L)(T)(R)xx(V)xxx(A)x{14}
EALGPVMSQHTGIDIGRKEGALGVFTACKLRSSVYHQAVALSPHNAIYR
1: EALGPVMSQHTGIDIGRKEGALGVFTACKLRSSVYHQAVALSPHNAIYR

1 YJX8_YEAST ck: 8234 len: 196 | P46989 saccharomyces cerevisiae (baker's
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
x{167}(L)x(S)(L)xx(F)(C)(K)xx(V)xxx(L)x{13}
MCGLTDVILPGKDAITTOIIDFNKNIFNVEETESALTTLKGAITWAGNSFPDAKLEPCQCNON
1: MCGLTDVILPGKDAITTOIIDFNKNIFNVEETESALTTLKGAITWAGNSFPDAKLEPCQCNON

1 YVSR11GRSTAOAGGYSAY ck: 1 len: 1 | P46989 saccharomyces cerevisiae (baker's
<(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W
x{167}(L)x(S)(L)xx(F)(C)(K)xx(V)xxx(L)x{13}
MCGLTDVILPGKDAITTOIIDFNKNIFNVEETESALTTLKGAITWAGNSFPDAKLEPCQCNON
1: MCGLTDVILPGKDAITTOIIDFNKNIFNVEETESALTTLKGAITWAGNSFPDAKLEPCQCNON

```

1 YKFE_ECOLI ck: 8141 len: 79 i P75677 escherichia coli. hypothetical 9.0 k
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{5}(L)x(P)(A)xx(F)(T)(R)xx(L)xxx(T)x{58}
MOSVLPFGGFFTRKQAOAVTTTISNLTLEDQSHERLVYVROTEGMVRANNEPDAEGEIN

1 YM27_MARPO ck: 7679 len: 69 i P38469 marchantia polymorpha (liverwort). h
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{20}(I)x(P)(A)xx(F)(S)(H)xx(P)xxx(G)x{33}
MASLSTFRTOTFLYPAHTFIYPATFHSFQPAFPGQALHFGGLDFERENAHCHIDNRKT

1 YN06_YEAST ck: 7226 len: 139 i P53842 saccharomyces cerevisiae (baker's ye
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{82}(A)x(S)(L)xx(L)(S)(K)xx(I)xxx(S)x{41}
x(L)x(S)(L)xx(I)(S)(K)xx(V)xxx(S)x{153}
MWLINHTYKLSTFLRKASNRFFNSSSSCSFLVFLVFVDFCSITSFLISFGLSSFLIF

1 YN01_CAEBL ck: 5825 len: 170 i P34496 caenorhabditis elegans. hypothetical 1
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x(L)x(S)(L)xx(I)(S)(K)xx(V)xxx(S)x{153}
MLRSLSTSKSTVACMSLTSKMAAEOPSKQEVDDLFABKRPQHNNPEQRHAYSVNKKVELGVALLD

1 YN09_CAEBL ck: 399 len: 336 i P34542 caenorhabditis elegans. hypothetical 1
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{11}(L)x(T)(I)x(L)(S)(R)xx(M)xxx(L)x{309}
MTLILFPASTLTKIILSLRIEMGRILEKRIIGTMSFKRIGNCSLGAIORLPNBSKLSSEKRR
GAGEFEFESKLASROARKTYITFKTKIDSEMIKTKKMLGVGENESIIVFLNMDR
VHRIPISKPKCPDHLQRILOMITTSEADQVLSKEPHPLRLQYVEKNEVFGFETLVPDHTLFLSDPFSRHPITTEGGGGGCPFM

1 YN06_CAEBL ck: 431 len: 381 i P50444 caenorhabditis elegans. hypothetical 1
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{115}(I)(I)x(P)(A)xx(L)(T)(R)xx(I)xxx(I)x{214}
x{115}(I)(I)x(P)(A)xx(L)(T)(R)xx(I)xxx(I)x{214}
MKQKRSIERGEGFYVLAEEBDMHITNLRIIDITIKASTIKRVSEISTGTTSORHMTLT
HEKGLKETEAVSTAMRHVYNLOVACVIVASGFKODFMQHLTAHADANKKKT
TEORAKFMILTHSSSGFKHAKLEVLETPVALRLADKAGEVKALNOFLMSTEPDRAFYGFHNVRANOEALIEITLVADSLFRAC

1 YORL_TTV1 ck: 9154 len: 232 i P19296 thermoproteus tenax virus 1 (strain
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{82}(V)x(P)(P)xx(I)(S)(K)xx(V)xxx(S)x{134}
MDVAVVILVSDDLTLTYLTLLHYMKRSQGLGQIYVSNINVLPRAKLIQCKRRGVGARGED
GOLIGDEGHIFIAKLISNRKNKNDMDFDITIKRKSIEEIEICEFORAL

1 YOTC_CAEBL ck: 9212 len: 161 i Q10120 caenorhabditis elegans. hypothetical 1
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{88}(A)x(T)(V)xx(M)(S)(H)xx(A)xxx(A)x{57}
MSRLFGIGYEGFDEMFVOOMIEKKSNAEOAKMLEOQKMKLECTETMPEESEPYPMKCLDFEBAF

1 YOXD_BACSU ck: 2171 len: 238 i P14802 bacillus subtilis. hypothetical oxid
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{115}(V)x(S)(A)xx(A)(S)(K)xx(V)xxx(I)x{72}
MOSLOHTALITGGGKRGATLALAKESGVNIGLGRTSANVEKVAEEVKAALGVKAFAADAVKX
IRVSALTLPSTVASDMSIELNLTGDEKPEKQEDAEIYVWOLKLPRIKTAGL
WSTNP

1 YPDC_ECOLI ck: 4986 len: 285 i P77396 escherichia coli. hypothetical trans
1: <(X){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x{115}(V)x(S)(A)xx(A)(S)(K)xx(V)xxx(I)x{72}
MOSLOHTALITGGGKRGATLALAKESGVNIGLGRTSANVEKVAEEVKAALGVKAFAADAVKX
IRVSALTLPSTVASDMSIELNLTGDEKPEKQEDAEIYVWOLKLPRIKTAGL
WSTNP

```

Mon Feb 14 08:07:43 2000

sp.cat

Page 25

RKLOPVPLOSNIKESAPSEKKEIKTNSFVLMNMTIMALMGVSYSP
TVSRFLGDEIGNEILITFCIGRGIGALVSKNGFNNNGLFTFTVLISIALFLFPFPAVSAIAALLFTIAMEYGLAKVKKQ
1 YVVP CAEEL ck: 1176 len: 274 : Q93834 caenorhabditis elegans. hypothetical
<(x){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x(19)(V)X(T)(L)X(X)(S)(R)X(X)(S)X(X)(S)X(249)
x(14)(L)(T)(T)X(X)(F)(C)(K)X(X)P)XXX(V)X(244)
1: MGSNVEOEKIDQVITFTFCCKWPTLAVGQVDPDHLTKSTNKIDGMELSLRTPYETPRKVV
AGAVVYVMTQTSIESMTPLALSPISMASKMKFGEQKGLVYRPERVSTNHR
VTAIKGEFEOKVMPDVNAPYHRLRMKPEVFTENRSELL
ZRP4_MAIZE ck: 6645 len: 364 : P47917 zea mays (maize). o-methyltransferase
<(x){1,200}(L,I,V,M,A,P)X(P,T,S)(L,I,V,M,A,P)XX(L,I,V,M,A,F,Y,W)(C
x(19)(V)X(T)(L)X(X)(S)(R)X(X)(S)X(X)(S)X(249)
MELSPNNSTQSLDQLEHMTFAPKSKMALKRAIHRADAIHLGGAASLSQILSKVHLPS
1: GLASDQIIVDAIKQSAEYFGISSLVAVGGIGAAQALSKAPPHKVCYDUL
HVAKAPLTHDVQFIAGDMFESIPADAVALLKSVLHMDHDCVILNCKKALPPBAGKVIITIMVYGAGPSDMKHEKMOALFDV
Databases searched:
SWISS-PROT, Release 38.1, Released on 20Nov1999, Formatted on 28Dec1999
Total finds: 337
Total length: 29,864,866
Total sequences: 82,229
CPU time: 03:13.70
1:AA_SEQUENCE 1.0 STANDARD: PRT: 133 AA.
ID AACP_PIG
AC Q09136
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 5'-AMP-ACTIVATED PROTEIN KINASE, GAMMA-1 SUBUNIT (AMPK GAMMA-1 CHAIN)
DE (38 KD SUBUNIT) (FRAGMENTS).
GN PRKAG1.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
NC Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN [1]
RP TISSUE=LIVER;
RX MEDLINE: 95050763.
RA STADLERON D., GAO G., MICHELL B.J., WIDMER J., MITCHELL K.,
RA TEH T., HOUSE C.M., WITTESS L.A., KEMP B.E.;
RT "Mammalian 5'-AMP-activated protein kinase non-catalytic subunits are
RT homologs of proteins that interact with yeast Snl protein kinase.";
RT J. Biol. Chem. 269:29343-29346(1994).
CC -1- FUNCTION: AMPK IS RESPONSIBLE FOR THE REGULATION OF FATTY ACID
CC SYNTHESIS BY PHOSPHORYLATION OF ACETYL-COA CARBOXYLASE. IT ALSO
CC REGULATES CHOLESTEROL SYNTHESIS VIA PHOSPHORYLATION AND
CC INACTIVATION OF HORMONE-SENSITIVE LIPASE AND
CC HYDROXYMETHYLGLUTARYL-COA REDUCTASE. APPEARS TO ACT AS A METABOLIC
CC STRESS-SENSING PROTEIN KINASE SWITCHING OFF BIOSYNTHETIC PATHWAYS
CC WHEN CELLULAR ATP LEVELS ARE DEPLETED AND WHEN 5'-AMP RISES IN
CC RESPONSE TO FUEL LIMITATION AND/OR HYPOXIA. THIS IS A REGULATORY
CC SUBUNIT.
CC -1- SUBUNIT. HETEROTRIMER OF AN ALPHA CATALYTIC SUBUNIT, A BETA AND A
CC GAMMA NON-CATALYTIC REGULATOR SUBUNITS.
CC -1- SIMILARITY: BELONGS TO THE 5'-AMP-ACTIVATED PROTEIN KINASE, GAMMA
CC SUBUNIT FAMILY.
CC -1- SIMILARITY: CONTAINS 4 CBS DOMAINS.
DR PFAM: PF00571; CBS 1.
KW Fatty acid biosynthesis; Repeat; CBS domain.
FT NON_TER 1
FT DOMAIN <1 >11 CBS 1.
FT DOMAIN <26 >42 CBS 2.
FT DOMAIN 62 91 CBS 3.
FT DOMAIN <95 127 CBS 4.
FT NON_CONS 11 12

FT NON_CONS 24 25
FT NON_CONS 42 43
FT NON_CONS 73 74
FT NON_CONS 80 81
FT NON_CONS 81 81
FT NON_CONS 94 95
FT NON_CONS 104 104
FT NON_CONS 103 104
FT NON_TER 133 133
SO SEQUENCE 133 AA: 14763 MW: 5D039F44 CRC32:
AACP_PIG Length: 133 February 14, 2000 08:02 Type: P Check: 8859
1 LVFDTSLQV KSAIVQIEL EHKXPVIDP ESGNTIYILF XLFITEPRK
51 PEFMSKSLKE LQIGTYANIA MVRVSAIPV IYKFDVINL AAEKSHYEG
101 VKLVYVDEN DVKGIYSLS DILQALVLTG GEX
1:AA_SEQUENCE 1.0 STANDARD: PRT: 77 AA.
ID ACP_BACSU
AC P80643; P51832.
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ACYL CARRIER PROTEIN (ACP).
GN ACPA OR ACP.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
CC Bacillus/staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE OF 15-77, AND SEQUENCE OF 1-14 FROM N.A.
RC STRAIN-168;
RX MEDLINE: 96326321.
RA MORBIDONT H.R., DE MENDOZA D., CRONAN J.E. JR.;
RT "Bacillus subtilis acyl carrier protein is encoded in a cluster of
RT lipid biosynthesis genes.";
RT J. Bacteriol. 178:4794-4800(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA OGURO A., KAKESHITA H., TAKAMATSU H., NAKAMURA K., YAMANE K.;
RA Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP MEDLINE: 94131947.
RX HEATON M.P., NEUBAUS F.C.;
RT "Role of the D-alanyl carrier protein in the biosynthesis of D-alanyl-
RT lipoteichoic acid.";
RT J. Bacteriol. 176:681-690(1994).
CC -1- FUNCTION: THIS PROTEIN IS THE CARRIER OF THE GROWING FATTY ACID
CC CHAIN IN FATTY ACID BIOSYNTHESIS.
CC -1- PATHWAY: KEY COMPONENT IN DE NOVO FATTY ACID BIOSYNTHESIS.
CC -1- PIM: THE GROWING FATTY ACID CHAINS ARE COVALENTLY BOUND TO THE
CC 4'-PHOSPHOPANTETHEINE PROSTHETIC GROUP.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
DR EMBL: D59433; AAC4308.1;
DR EMBL: D64116; BAA10975.1;
DR EMBL: Z98112; CAB13465.1;
DR HSRP: P02901; IACP.
DR SUBUNIT1; HG11536; ACPA.
DR PROSITE: PS0012; PHOSPHOPANTETHEINE. 1.
DR PROSITE: PS0075; ACP DOMAIN; 1.
DR PFAM: PF00570; PP-binding; 1.
KW Fatty acid biosynthesis; Phosphopantetheine.
FT BINDING 77 AA: 8591 MW: 48D5166 CRC32:
SO SEQUENCE

ACP_BACSU Length: 77 February 14, 2000 08:02 Type: P Check: 8148 ..

1 MADLELVK IYDRIGVDE ADVKLEASF EDGASLDV VELVLELDE
51 FMEISPEDA EKIAVGDV NYIONQ

11AA_SEQUENCE 1.0 STANDARD; PRT; 253 AA.
ID ADH2_DROMO
AC P09369;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE ALCOHOL DEHYDROGENASE 2 (EC 1.1.1.1).
GN ADH2.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pelegronota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NICHOLS;
RA MEDLINE; 8937903.
RT ABRINKSON P.W., MILLS L.E., STARMER W.T., SULLIVAN D.T.;
RT "Structure and evolution of the Adh genes of Drosophila melanogaster";
RT Genetics 120:713-723(1988).
CC -1- CATALYTIC ACTIVITY: ALCOHOL + NAD(+) = ALDEHYDE OR KETONE + NADH.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC FAMILY (SDS).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M37276; AAA28334.1; -
DR EMBL: X12536; CAA31054.1; -
DR PIR: S01901; S01901
DR PROSITE: PS00061; ADH_SHORT; 1.
DR PFAM: PF00106; adh_short; 1.
DR PFM: PF00663; adh_short; 1.
KW Oxidoreductase; NAD.
FT INIT_MET 0
FT NP_BIND 9 32 NAD (BY SIMILARITY).
FT ACT_SITE 150 150 BY SIMILARITY.
SQ SEQUENCE 253 AA; 27382 MW; 0136EC76 CRC32;
ADH2_DROMO Length: 253 February 14, 2000 08:02 Type: P Check: 8528 ..
1 AIAKNNITVF AGIGIGFDT SREIYKSGPK NLVILDRLEN PAIAELKAL
51 NPKYVTFYP YDVYTSVAET TKLKTIFDK LKTVDLLING TGILDDHOIE
101 RTIAVNTGT LNTTAINSF WDKRKGPGG VIANICVYG FNAIIPVYV
151 SASKAALSF TNSIARLAPI TGVTAYSNP GIKRTLVHK FNSMIDVEPR
201 VAEILLEHPT OTTICQANF VKAIOANNG AIAKIDGLT EAIETKIMD
251 SHI
11AA_SEQUENCE 1.0 STANDARD; PRT; 357 AA.
ID ALR_TREPA
AC Q56346;
DT 15-DEC-1999 (Rel. 39, Created)
DT 15-DEC-1999 (Rel. 39, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE ALANINE RACEMASE (EC 5.1.1.1).

GN ALR OR TP0681.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NICHOLS;
RA MEDLINE; 98332770.
RA FRASER C.M., NORKIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.,
RA DODSON R., GWINN M., HICKEY E.K., CLAYTON R., KETCHUM R.A.,
RA SODERGREN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,
RA KHALK H., RICHARDSON D., HOWELL J.K., CHIDAMBARAM M., UTTERBACK T.,
RA MCDONALD L., ARTACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,
RA HATCH B., HORST K., ROBERTS K., MATTHEY L., WEIDMAN J., SMITH H.O.,
RA VENTER J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete";
RT Science 281:375-388(1998).
RL Science 281:375-388(1998).
RN [2]
RP SEQUENCE OF 1-178 FROM N.A.
RC STRAIN-NICHOLS;
RA STEINER B.M., RODES B.;
RT Partial sequence of alanine racemase from Treponema pallidum";
RT Submitted (May 1996) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: L-ALANINE - D-ALANINE.
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -1- SIMILARITY: BELONGS TO THE ALANINE RACEMASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001242; AAC65644.1; -
DR EMBL: U57756; AAB17466.1; -
DR TIGR: TP0681; -
DR PROSITE: PS00395; ALANINE_RACEMASE; FALSE_NEG.
DR PFAM: PF00842; Ala_racemase; 1.
KW Isomerase; Pyridoxal phosphate.
FT BINDING 13 13 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT SEQUENCE 357 AA; 38098 MW; 88BD7754 CRC32;
ALR_TREPA Length: 357 February 14, 2000 08:02 Type: P Check: 255 ..
1 MSRTARVCL PVKADAYGHG ACDAQAALS CGVSHFAVAC VOEASOLRAA
51 GVRAPILCLST TPAEELISL IEHRVHTVIS ERAHIALIAR ALROSADTGA
101 TCGVHWKIDP GNGRIGCAPD EACALVOMVC APFGHLEGV CTRHSVADSV
151 RAEDLQTEM QRAHFMHCVO YIKSGISIP LVHANSAL LCHRAHFDK
201 VRGLIATGY APESVPAVR SVFLPWVLY TVYRAIKIP AGAVYSIQRL
251 WRAHTEHWG ILPIGYADGV MRALSPGLOV CIGKWPYVY GAICMOCVY
301 DLGTPLRVTV GDRVTLFSPQ DAGGPGOGAD VLAHAGTIP YELCAIGKR
351 VERYIIR
11AA_SEQUENCE 1.0 STANDARD; PRT; 289 AA.
ID AMIA_SALTY
AC P33772;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PROBABLE N-ACETYLURAMOYL-L-ALANINE AMIDASE AMIA PRECURSOR
DE (EC 3.5.1.28).
GN AMIA.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

CC Salmoneella.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LT2;
 RX MEDLINE: 93352403.
 RA XU K., ELLIOTT T.;
 RT "An oxygen-dependent coproporphyrinogen oxidase encoded by the hamf
 U gene of Salmoneella typhimurium.";
 RU Bacteriol. 175:4990-4999(1993).
 CC -1- NUCLEOTIC ACTIVITY: HYDROLYZES THE LINK BETWEEN N-ACETYLMURAMOYL
 RESIDUES AND L-AMINO ACID RESIDUES IN CERTAIN BACTERIAL CELL-WALL
 GLYCOPOLYMERES.
 CC -1- SIMILARITY: STRONG, TO B. SUBTILIS LYTIC/CWLB.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L19503; AAA27138.1; -
 DR STRYGENE; SG10008; AMIA.
 DR PFAM: PF01520; Amidase_3; 1.
 KW Hydrolyase; Cell wall; Signal.
 FT SIGNAL 1 34 POTENTIAL.
 FT CHAIN 35 289 PROBABLE N-ACETYLMURAMOYL-L-ALANINE
 FT AMIDASE AMIA.
 FT SEQUENCE 289 AA; 31659 MW; 5F107FEC CRC32;
 SO AMIA, SALTY Length: 289 February 14, 2000 08:02 Type: P Check: 5663 ..

1 MSTFLUKTL TSKROYLKTG IALTLGSKS HAVAKETTLK TNGSHKPKT
 51 KITSKRLVM LDPHGIGIDT GAIGNGSGE KHYVLATLAKN VVALRNHGI
 101 DARLTRIGDT FIPLYDRVEI AKHGDADLFM SIHDSGTNP KAAGASVFL
 151 SNRGAASSMA KYLSERENRA DEVAKRKATD RDHLQGVLF DVGQDTLKN
 201 STLLGSHLK KIKPIHKLHS RTTEQAAPV LKSPISYVL VELSFTINPE
 251 EERLLGTATF ROKITAIAN GIISTFHFMD NOKAHTKKR

11AA_SEQUENCE 1.0
 ID AMPC_SERMA STANDARD: PRT: 376 AA.
 AC P18539;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE BETA-LACTAMASE PRECURSOR (EC 3.5.2.6) (CEPHALOSPORINASE).
 GN AMPC.
 OS Serratia marcescens.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Serratia.
 CC [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SR50;
 RX MEDLINE: 91032898.
 RA NOMURA K., YOSHIDA T.;
 RT "Nucleotide sequence of the Serratia marcescens SR50 chromosomal ampc
 beta-lactamase gene.";
 RU FEMS Microbiol. Lett. 58:295-299(1990).
 CC -1- FUNCTION: THIS PROTEIN IS A SERINE BETA-LACTAMASE WITH A SUBSTRATE
 CC SPECIFICITY FOR CEPHALOSPORINS.
 CC -1- CATALYTIC ACTIVITY: A BETA-LACTAM + H(2)O -> A SUBSTITUTED BETA-
 CC AMINO ACID.
 CC -1- SUBCELLULAR LOCATION: PERIPLASMIC (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE CLASS C BETA-LACTAMASE FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X52964; CA37137.1; -
 DR PIR: S11710; QKSE.7.
 DR PIR: A48176; A48176.
 DR HSSP: P05364; 1BL5.
 DR PROSITE: PS00336; BETA-LACTAMASE_C_1.
 DR PFAM: PF00144; beta-lactamase; 1.
 KW Hydrolyase; Antibiotic resistance; Periplasmic; Signal.
 FT SIGNAL 1 21
 FT CHAIN 22 376
 FT ACT_SITE 79 79 BY SIMILARITY.
 FT BINDING 328 330 SUBSTRATE (BY SIMILARITY).
 FT SEQUENCE 376 AA; 41096 MW; 016BEF00 CRC32;
 SO AMPC_SERMA Length: 376 February 14, 2000 08:02 Type: P Check: 107 ..

1 MTKMRCAL IALILPTAH AAQODIDAV IOPLMKRYGV PGMAINVSVD
 51 GKQQLPYGV ASKOTGKPI EQLTFEVSGL SKTFATLAV YAOOSKLSF
 101 KDPASHYLD VRGSAFDGVS LNLATHTSG LPLFVPDDV NNAOLMAYVR
 151 AMQPKHPAG YRYVSNLIGS MGMTAKSL DQPIQMEQ GMLPALGSH
 201 TYQVPAAM ANVAGYSKD DKPVVNP GP LDASVGIKS NNRDLIRYLD
 251 ANIQYKVAS VARKPRPTS VITSAGFTQ DLMENVPY VKLSRIEEN
 301 NAGMNGTTP ATAITPPOPE LRAGWNTKG STGFSFIYAV FIPKANIAYE
 351 MANKWFPND DVEPAVHII QALEKR

11AA_SEQUENCE 1.0
 ID AMPC_CITFR STANDARD: PRT: 187 AA.
 AC Q00831;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE AMPD PROTEIN.
 GN AMPD.
 OS Citrobacter freundii, and Enterobacter cloacae.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Citrobacter.
 CC [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OS60, AND 14;
 RX MEDLINE: 9319292.
 RA KOPP U., KIEDEMAN B., LINDQUIST S., NORMARK S.;
 RT "Sequences of wild-type and mutant ampd genes of Citrobacter freundii
 and Enterobacter cloacae.";
 RU Antimicrob. Agents Chemother. 37:224-228(1993).
 CC -1- FUNCTION: PUTATIVE SIGNALING PROTEIN IN BETA-LACTAMASE
 CC REGULATION. AMPD SEEMS NOT TO ACT AS A DIRECT SENSOR FOR
 CC BETA-LACTAMS.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z14002; CA478390.1; -
 DR EMBL: Z14003; CA478391.1; -
 DR PIR: A48901; A48901.
 DR PIR: S26138; S26138.

DR PIR: S26139; S26139.
 SO SEQUENCE 187 AA; 20839 MW; 229DF9F2 CRC32;
 AMPD_CITER Length: 187 February 14, 2000 08:02 Type: P Check: 4831 ..

1 MLENGWLVD ARHVSPHHD CRPEDEKPTL LVYHNISLPP GEGGPMWIDA
 51 LETGTIDPA HPEFAEIAL ALSDACLIR DGEVGYVPE DKRMHAGVS
 101 MYGRERCND FSGIELEGT DTPPYTDAQ EKLVAVTQL IGRPAIDN
 151 ITGSHDAPE RKTDEPAFD WSRFHMLTT SSDKEIT

11AA_SEQUENCE 1.0 STANDARD; PRT: 339 AA.
 ID ANX2_XENIA
 AC P24801;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last annotation update)
 DE ANNEXIN II TYPE II (LIPOCORTIN II) (CALPACTIN I HEAVY CHAIN)
 DE (CHROMOBINDIN 8) (P36) (PROTEIN I) (PLACENTAL ANTICOAGULANT PROTEIN
 DE IV) (PAP-IV).
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
 OC Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopodinae;
 OC Xenopus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-OOCYTE;
 RX MEDLINE: 92011609.
 RA IZANT J.G.; BRYSON L.J.;
 RT "Xenopus annexin II (calpactin I) heavy chain has a distinct amino
 RT terminus".
 RL J Biol. Chem. 266:18560-18566(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-KIDNEY;
 RX MEDLINE: 92009222.
 RA GERKE V.; KOCH W.; THIEL C.;
 RT Primary structure and expression of the Xenopus laevis gene encoding
 RT annexin II".
 RL Gene 104:259-264(1991).
 CC -1- FUNCTION: CALCIUM-REGULATED MEMBRANE-BINDING PROTEIN WHOSE
 CC AFFINITY FOR CALCIUM IS GREATLY ENHANCED BY ANIONIC PHOSPHOLIPIDS.
 CC -1- BINDS TWO CALCIUM IONS WITH HIGH AFFINITY.
 CC -1- SUBUNIT: Tetramer of 2 light chains (P10 proteins) and 2 heavy
 CC chains (P36 proteins).
 CC -1- SUBCELLULAR LOCATION: IN THE LAMINA BENEATH THE PLASMA MEMBRANE.
 CC -1- TISSUE SPECIFICITY: ADULT BRAIN, HEART, STRIATED MUSCLE, LIVER,
 CC KIDNEY, AND VERY HIGH LEVELS IN SKIN.
 CC -1- DEVELOPMENTAL STAGE: THROUGHOUT OOGENESIS AND IN MATURE EGGS.
 CC CONSTANT LEVELS DURING EARLY EMBRYOGENESIS, BUT DECREASE AT 8H
 CC AFTER MIDLASTULA TRANSITION, THE STEADY STATE LEVEL INCREASES
 CC SUBSTANTIALLY.
 CC -1- DOMAIN: CONTAINS FOUR HOMOLOGOUS REPEATS WITH A CONSENSUS
 CC SEQUENCE COMMON TO ALL ANNEXIN PROTEINS. A PAIR OF THESE REPEATS
 CC MAY FORM ONE BINDING SITE FOR CALCIUM AND PHOSPHOLIPID.
 CC -1- MISCELLANEOUS: IT MAY CROSS-LINK PLASMA MEMBRANE PHOSPHOLIPIDS
 CC WITH ACTIN AND THE CYTOSKELETON AND BE INVOLVED WITH EXOCYTOSIS.
 CC -1- SIMILARITY: BELONGS TO THE ANNEXIN FAMILY.
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC
 CC EMBL: M58575; AAA49664.1; -
 CC EMBL: M58575; AAA49665.1; -
 CC EMBL: M60769; AAA49886.1; -
 CC PIR: J01298; J01298.

DR PIR: A41002; A41002.
 DR PIR: B41002; B41002.
 DR HSSE; P04083; 1809.
 DR PROSITE; PS00223; ANNEXIN; 4.
 DR PFAM; PF00191; annexin; 4.
 KW Annexin; Calcium/phospholipid-binding; Repeat; Phosphorylation.
 FT INIT MET 0
 FT DOMAIN 1 24
 FT REPEAT 42 102
 FT REPEAT 114 174
 FT REPEAT 199 259
 FT REPEAT 274 334
 FT MOD_RES 26 26
 FT VARIANT 243 243
 FT VARIANT 257 257
 FT CONFLICT 222 223
 FT CONFLICT 297 297
 FT SEQUENCE 339 AA; 38643 MW; 01CF92AC CRC32;
 SO SEQUENCE

ANX2_XENIA Length: 339 February 14, 2000 08:02 Type: P Check: 6452 ..

1 ALIHEILKRL SLGNOSSR OSKIGSVRAA THPDAEKDAA ALETAIKTG
 51 VDELTIIINIL TNRSNEOROD IAFAPRRRK KDLPSALKGA LSGNLETVM
 101 GLIKTRPOYD ASELKASMKG LGTDEDLIE IICSTNKEL LDIONAVREL
 151 FTELEKDIM SDTSGDEPKL MYALAKGRQ EDGNMVDYER IDQARELYE
 201 AGVRRKGTDV TKWTITMER SHPHLOKVEE RYKSYSPYDI ESIKKEVKG
 251 DENNFMLVY OCIONKPLYE ADRLVESMKG KGTEDKILIR IWSNRNLDW
 301 LKIRQEFKKK YKSLHFFIG QDTKGDYORA LNLGSGSD

11AA_SEQUENCE 1.0 STANDARD; PRT: 360 AA.
 ID ARG1_XENIA
 AC Q91531;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ARGINASE, NON-HEPATIC I (EC 3.5.3.1).
 GN ARG1.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
 OC Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopodinae;
 OC Xenopus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-INTESTINE;
 RX MEDLINE: 95014323.
 RA PATTERSON D.; SHI Y.-B.;
 RT "thyroid hormone-dependent differential regulation of multiple
 RT arginase genes during amphibian metamorphosis".
 RL J. Biol. Chem. 269:25328-25334(1994).
 CC -1- FUNCTION: AS WELL AS ITS ROLE IN THE UREA CYCLE, MAY BE INVOLVED
 CC IN TISSUE REMODELING.
 CC -1- CATALYTIC ACTIVITY: L-ARGININE + H(2)O = L-ORNITHINE + UREA.
 CC -1- COFACTOR: MANGANESE (BY SIMILARITY).
 CC -1- PATHWAY: FIRST STEP IN ARGININE DEGRADATION.
 CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT DIFFERING TADPOLE STAGES IN TAIL,
 CC INTESTINE, HINDLIMB AND TRUNK REGION. MOST ABUNDANT IN TADPOLE
 CC TAIL.
 CC -1- DEVELOPMENTAL STAGE: FIRST DETECTED IN NEURULA (STAGE 16/17).
 CC HIGHEST LEVELS IN WHOLE TADPOLE FOUND AROUND STAGE 47/48. IN THE
 CC INTESTINE, INCREASED LEVELS ARE FOUND DURING METAMORPHOSIS (STAGES
 CC 58-64). LOW LEVELS EXPRESSED IN HINDLIMB UNTIL STAGE 66 AFTER
 CC WHICH, LEVELS DRAMATICALLY INCREASE. IN THE TAIL, A CONSTANT HIGH
 CC LEVEL OF EXPRESSION IS FOUND THROUGHOUT METAMORPHOSIS.
 CC -1- INDUCTION: ACTIVATED BY THYROID HORMONE (T3).
 CC -1- SIMILARITY: BELONGS TO THE ARGINASE FAMILY.
 CC
 CC
 CC

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: 008406; AA:56891.1; ..
 DR HSBP; P07824; 2RLA.
 DR PROSITE; PS00147; ARGINASE_1; 1.
 DR PROSITE; PS00148; ARGINASE_2; 1.
 DR PROSITE; PS01053; ARGINASE_3; 1.
 DR PFM; PF00491; arginase; 1.
 DR Urea cycle; Arginine metabolism; Hydrolase; Manganese;
 KW Multigene family.
 FT METAL 122 122 MANGANESE 1 (BY SIMILARITY).
 FT METAL 145 145 MANGANESE 1 AND 2 (BY SIMILARITY).
 FT METAL 147 147 MANGANESE 2 (BY SIMILARITY).
 FT METAL 149 149 MANGANESE 1 (BY SIMILARITY).
 FT METAL 253 253 MANGANESE 1 AND 2 (BY SIMILARITY).
 FT METAL 255 255 MANGANESE 2 (BY SIMILARITY).
 SQ SEQUENCE 360 AA; 39155 MW; 31E02C27 CRC32;

ARG1_XENLA Length: 360 February 14, 2000 08:02 Type: P Check: 4321 ..

1 MSRSNPFVRL LKKOVSIIKL OKKCSHSVAV IGAPFSKGOK RRGVEHGPAA
 51 IRSAGLIERL SNLGCNVCDF GDHFSGQVFN DELYSIVKH PRVGLACKV
 101 LAEVSXAVG AGHTCVTLGG DSHLAFGSIT GHAQCPDLC VIWDAHADI
 151 NPLUTPSSGN LHGPVSFLL RELQDKVPPI PGFSMAKPL SKSDIYIGL
 201 RDLDPAEPTI LKNYDISYS MRHIDCMGK KYMEKTFDOL LGRDRPHIL
 251 SEDIDAFDPA LAPATGTPVI GGLTYREGYV ITEEIHNTGM LSAVDLYEVN
 301 PVLATSEEV KATANLADV IASCGQTR GAHTRADTII DYLPTSTSY
 351 ESDNEQVRI

11AA_SEQUENCE 1.0 STANDARD; PRT; 360 AA.
 ID ARG2_XENLA
 AC Q91554;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE ARGINASE, NON-HEPATIC 2 (EC 3.5.3.1).
 GN ARG2.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
 OC Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopodinae;
 OC Xenopus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-INTESTINE;
 RX MEDLINE: 95014323.
 RA PATTERSON D., SHI Y.-B.;
 RT "Thyroid hormone-dependent differential regulation of multiple
 RT arginase genes during amphibian metamorphosis.";
 RL J Biol Chem 269:25328-25334(1994)
 CC -1- FUNCTION: AS WELL AS ITS ROLE IN THE UREA CYCLE, MAY BE INVOLVED
 CC -1- IN TISSUE REMODELING.
 CC -1- CATALYTIC ACTIVITY: L-ARGININE + H(2)O = L-ORNITHINE + UREA.
 CC -1- COFACTOR: MANGANESE (BY SIMILARITY).
 CC -1- PATHWAY: FIRST STEP IN ARGININE DEGRADATION.
 CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT DIFFERING TADPOLE STAGES IN
 CC TAIL, INTESTINE, HINDLIMB AND TRUNK REGION. STRONGEST IN TADPOLE
 CC TAIL.
 CC -1- DEVELOPMENTAL STAGE: FIRST DETECTED IN NEURULA (STAGE 16/17).
 CC HIGHEST LEVELS IN WHOLE TADPOLE FOUND AROUND STAGE 47/48. IN THE

INTESTINE, INCREASED LEVELS ARE FOUND DURING METAMORPHOSIS (STAGES 58-64) AND IN THE HINDLIMB, EXPRESSED AT LOW LEVELS DURING METAMORPHOSIS UNTIL STAGE 66 WHEN LEVELS DRAMATICALLY INCREASE. IN THE TAIL, A CONSTANT HIGH LEVEL OF EXPRESSION IS FOUND THROUGHOUT METAMORPHOSIS.
 CC -1- INDUCTION: ACTIVATED BY THYROID HORMONE (T3).
 CC -1- SIMILARITY: BELONGS TO THE ARGINASE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: 008407; AA:56892.1; ..
 DR HSBP; P07824; 2RLA.
 DR PROSITE; PS00147; ARGINASE_1; 1.
 DR PROSITE; PS00148; ARGINASE_2; 1.
 DR PROSITE; PS01053; ARGINASE_3; 1.
 DR PFM; PF00491; arginase; 1.
 DR Urea cycle; Arginine metabolism; Hydrolase; Manganese;
 KW Multigene family.
 FT METAL 122 122 MANGANESE 1 (BY SIMILARITY).
 FT METAL 145 145 MANGANESE 1 AND 2 (BY SIMILARITY).
 FT METAL 147 147 MANGANESE 2 (BY SIMILARITY).
 FT METAL 149 149 MANGANESE 1 (BY SIMILARITY).
 FT METAL 253 253 MANGANESE 1 AND 2 (BY SIMILARITY).
 FT METAL 255 255 MANGANESE 2 (BY SIMILARITY).
 SQ SEQUENCE 360 AA; 39154 MW; 26CE48BC CRC32;

ARG2_XENLA Length: 360 February 14, 2000 08:02 Type: P Check: 4641 ..

1 MSRSNPFVRL LKKOVSIIKL OKKCSHSVAV IGAPFSKGOK RRGVEHGPAA
 51 IRSAGLIERL SNLGCNVCDF GDHFSGQVFN DELYSIVKH PRVGLACKV
 101 LAEVSXAVG AGHTCVTLGG DSHLAFGSIT GHAQCPDLC VIWDAHADI
 151 NPLUTPSSGN LHGPVSFLL RELQDKVPPI PGFSMAKPL SKSDIYIGL
 201 RDLDPAEPTI LKNYDISYS MRHIDCMGK KYMEKTFDOL LGRDRPHIL
 251 SEDIDAFDPA LAPATGTPVI GGLTYREGYV ITEEIHNTGM LSAVDLYEVN
 301 PVLATSEEV KATANLADV IASCGQTR GAHTRADTII DYLPTSTSY
 351 ESDNEQVRI

11AA_SEQUENCE 1.0 STANDARD; PRT; 360 AA.
 ID ARG3_XENLA
 AC Q91555;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE ARGINASE, NON-HEPATIC 3 (EC 3.5.3.1).
 GN ARG3.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
 OC Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopodinae;
 OC Xenopus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-INTESTINE;
 RX MEDLINE: 95014323.
 RA PATTERSON D., SHI Y.-B.;
 RT "Thyroid hormone-dependent differential regulation of multiple
 RT arginase genes during amphibian metamorphosis.";
 RL J Biol Chem 269:25328-25334(1994)
 CC -1- FUNCTION: AS WELL AS ITS ROLE IN THE UREA CYCLE, MAY BE INVOLVED
 CC -1- IN TISSUE REMODELING.
 CC -1- CATALYTIC ACTIVITY: L-ARGININE + H(2)O = L-ORNITHINE + UREA.

CC -1- COFACTOR: MANGANESE (BY SIMILARITY).
 CC -1- PATHWAY: FIRST STEP IN ARGININE DEGRADATION.
 CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT DIFFERING TADPOLE STAGES IN
 CC TAIL, INTESTINE, HINDLIMB AND TRUNK REGION. STRONGEST IN TADPOLE
 CC TAIL.
 CC -1- DEVELOPMENTAL STAGE: FIRST DETECTED IN EARLY TAILBUD (STAGE
 CC 23/24). HIGHEST LEVELS IN WHOLE TADPOLE FOUND AROUND STAGE 47/48.
 CC IN THE INTESTINE. INCREASED LEVELS ARE FOUND DURING METAMORPHOSIS
 CC (STAGES 58-64) AND IN THE HINDLIMB, EXPRESSED AT LOW LEVELS DURING
 CC METAMORPHOSIS UNTIL STAGE 66 WHEN LEVELS DRAAMATICALLY INCREASE. IN
 CC THE TAIL, A CONSTANT HIGH LEVEL OF EXPRESSION IS FOUND THROUGHOUT
 CC METAMORPHOSIS.
 CC -1- INDUCTION: ACTIVATED BY THYROID HORMONE (T3).
 CC -1- SIMILARITY: BELONGS TO THE ARGINASE FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U08408; AAA56893.1; -
 CC HSSP: P07824; 2RLA.
 CC PROSITE: PS00147; ARGINASE_1; 1.
 CC PROSITE: PS00148; ARGINASE_2; 1.
 CC PROSITE: PS01053; ARGINASE_3; 1.
 CC PRAM: PF00491; arginase; 1.
 CC Urea cycle: Arginine metabolism; Hydrolase; Manganese;
 CC Multigene family.
 CC METAL 122 122 MANGANESE 1 (BY SIMILARITY).
 CC METAL 145 145 MANGANESE 1 AND 2 (BY SIMILARITY).
 CC METAL 147 147 MANGANESE 2 (BY SIMILARITY).
 CC METAL 149 149 MANGANESE 1 (BY SIMILARITY).
 CC METAL 253 253 MANGANESE 1 AND 2 (BY SIMILARITY).
 CC METAL 255 255 MANGANESE 2 (BY SIMILARITY).
 CC SEQUENCE 360 AA; 39238 MW; A89683FA CRC32;
 SQ
 ARG3_XENLA Length: 360 February 14, 2000 08:02 Type: P Check: 4724 ..
 1 MSIRSNFVRL LKKQVNIKL QKKCSHSVAV IGAPFSKQK RRGVHSPAA
 51 IRSAGLIDRL SNLGCNCFD GDHFSGVPM DEQNSIYKH PRVGLACKV
 101 LAKEVGNVAG AGHTCVTLGG DSHLAFCSIT GHQCCPDLC VIWDAHADI
 151 NPLTTPSGN LHGPVPSFL RELODKRTPP PGFSNAKCL SKSDIYIGL
 201 RDDPAPFOFI LKNYISYYS MRHIDCKGR KYMETFPOL LGRDRRIHL
 251 SPIDAPDPA LAPATGTPVI GGLTYRGGV ITTEHNHGM LSLADLEVN
 301 PVLTATSEEV KATANLAVDV IASCGOTRE GAHTRADITI DVLPFTSTY
 351 ESDNEOVRI
 11A_SEQUENCE 1.0
 ID ARGC_ECOLI STANDARD; PRT; 334 AA.
 AC P11446;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, last sequence update)
 DT 15-DEC-1998 (Rel. 37, last annotation update)
 DE N-ACETYL-GAMMA-GLUTAMYL-PHOSPHATE REDUCTASE (EC 1.2.1.38) (AGPR) (N-
 DE ACETYL-L-GLUTAMATE SEMIALDEHYDE DEHYDROGENASE) (NAGSA DEHYDROGENASE).
 GN ARGC
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 89121510.

RA PARSON C., BOYEN A., COHEN G.N., GLANDORFF N.:
 RT "Nucleotide sequence of Escherichia coli argB and argC genes:
 RT comparison of N-acetylglutamate kinase and
 RT N-acetylglutamate-gamma-semialdehyde dehydrogenase with homologous
 RT and analogous enzymes.";
 RT Gene 68:275-283(1988).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN-K12 / MG1655;
 RC MEDLINE: 94089392.
 RX BLATTER F.R., BURLAND V.D., PLUNKETT G. III, SOFIA H.J.,
 RA DANIELS D.L.:
 RT "Analysis of the Escherichia coli genome. IV. DNA sequence of the
 RT region from 89.2 to 97.8 minutes.";
 RT Nucleic Acids Res. 21:5408-5417(1993).
 RL [3]
 RN SEQUENCE OF 1-48 FROM N.A.
 RP STRAIN-K12;
 RC MEDLINE: 83143275.
 RX PIETTE J., CUNIN R., BOYEN A., CHARLIER D.R.M., CRABEEL M.,
 RA VAN VLIET F., GLANDORFF N., SOUIRES C., SOUIRES C.L.:
 RT "The regulatory region of the divergent argECB operon in Escherichia
 RT coli K-12.";
 RL Nucleic Acids Res. 10:8031-8048(1982).
 RN [4]
 RP SEQUENCE OF 1-19 FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE: 92202162.
 RA MEINDEL T., SCHWITZ E., MECHULAM Y., BLANQUET S.:
 RT "Structural and biochemical characterization of the Escherichia coli
 RT argE gene product.";
 RT J. Bacteriol. 174:2323-2331(1992).
 CC -1- CATALYTIC ACTIVITY: N-ACETYL-L-GLUTAMATE-5-SEMIALDEHYDE + NADP(+)
 CC + ORTHOPHOSPHATE -> N-ACETYL-5-GLUTAMYL-PHOSPHATE + NADPH.
 CC -1- PATHWAY: THIRD STEP IN ARGININE BIOSYNTHESIS.
 CC -1- SIMILARITY: TO OTHER BACTERIAL ARGC AND TO NAGSA DOMAIN OF FUNGAL
 CC PROTEIN ARG3/6.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M21446; AAA23477.1; -
 CC EMBL: J01587; AAB59146.1; -
 CC EMBL: X55417; -; NOT ANNOTATED_CDS.
 CC EMBL: U00006; AAC43064.1; -
 CC EMBL: AE000470; AAC76840.1; -
 CC PIR: J0332; RDECEP.
 CC ECOSYSTEM: B610065; ARGC.
 CC PROSITE: PS01224; ARGC; 1.
 CC PRAM: PF01118; Semialdehyde dh; 1.
 CC Arginine biosynthesis; Oxidoreductase; NADP.
 CC ACT_SITE 154 154 BY SIMILARITY.
 CC SEQUENCE 334 AA; 35952 MW; DB98AB60 CRC32;
 SQ
 ARGC_ECOLI Length: 334 February 14, 2000 08:02 Type: P Check: 6342 ..
 1 MLNLIIVGAS GYAGALVTV VNRHPHNIIT ALTVASQNSD AGKLISDLHP
 51 QLKGIIVDPL QPKSDISEFS PGVNVFLAT AHEVSHDLAP QLEAGCVVF
 101 DLSGAFRND ATPEKYIYGF THQYPELLEQ AAYGLAEMWG NKLEENANLA
 151 VPGETPRAAQ LAKPLRIDAD LLDLNQNPVI NATSGVSGAG RKAALNSFC
 201 EYSLQPTGVF TRHROPEIAT HLGADVIFTP HLGNEFRGIL ETITCRKSG
 251 VTAQVAQVL QQAYAHKPLV RLYDKGPAL KNYVGLPFCD IGFAVQGBHL

301 IIVATEDNILL KGAQAQAVOC ANIRFGYAET QSLI

11AA_SEQUENCE 1.0 STANDARD; PRT: 142 AA.

ATP8_HUMAN

AC P48201; MEDLINE: 95213032.

DR 01-FEB-1996 (Rel. 33, Created)

DR 01-FEB-1996 (Rel. 33, Last sequence update)

DR 15-DEC-1998 (Rel. 37, Last annotation update)

DE ATP SYNTHASE LIPID-BINDING PROTEIN P3 PRECURSOR (EC 3.6.1.34) (ATPASE PROTEIN 9) (SUBUNIT C).

GN ATP8.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eukarya; Primates; Catarrhini; Homnidae; Homo.

CC [1]

CC SEQUENCE FROM N.A.

CC TISSUE=LIVER;

CC MEDLINE: 95213032.

CC YAN W. L., LERNER T. J., HAINES J. L., GUSELLA J. F.;

CC "Sequence analysis and mapping of a novel human mitochondrial ATP synthase subunit 9 cDNA (ATP8G3).";

CC Genomics 24:375-377(1994).

CC -1- FUNCTION: THIS PROTEIN IS ONE OF THE CHAINS OF THE NONENZYMATIC MEMBRANE COMPONENT (F0) OF MITOCHONDRIAL ATPASE.

CC SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0) HAS THREE MAIN SUBUNITS: A, B AND C.

CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL MEMBRANE.

CC -1- DISEASE: THIS PROTEIN IS THE MAJOR PROTEIN STORED IN THE STORAGE BODIES OF ANIMALS OR HUMAN AFFECTED WITH CEROID LIPOFUSCINOSIS (BATTEN'S DISEASE).

CC -1- MICELLAR PROPERTIES: THERE ARE THREE GENES WHICH ENCODE THE MITOCHONDRIAL ATP SYNTHASE PROTEOLIPID AND THEY SPECIFY PRECURSORS WITH DIFFERENT IMPORT SEQUENCES BUT IDENTICAL MATURE PROTEINS.

CC -1- SIMILARITY: BELONGS TO THE ATPASE C CHAIN FAMILY.

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL: U09813; AAA78807.1; -

CC MIM: 602736; -

CC DR PROSITE: PS00605; ATPASE_C.1.

CC DR PFAM: PF00137; ATP-synt_C.1.

CC KW Hydrogen ion transport; Lipid-binding; Mitochondrion; CF(0); Transmembrane; Transmembrane; Transmembrane; Transmembrane.

CC FT TRANSIT 1 67 MITOCHONDRION.

CC CHAIN 68 142 ATP SYNTHASE LIPID-BINDING PROTEIN.

CC SEQUENCE 142 AA; 14693 MW; 70215F04 CRC32;

ATP8_HUMAN Length: 142 February 14, 2000 08:02 Type: P Check: 2178

1 MRACALACT PSURAGSRV AYRPAISVY SRPEASRGE GSTVFNAGN.

51 GYSQIQREF QTSASRID TAKFTGAGA ATVGAGGA GIGIVGSLI

101 IGYARNPSLK QQLSYALIG FALSEAMGLF CLMAFLILF AM

11AA_SEQUENCE 1.0 STANDARD; PRT: 50 AA.

ATP8_PODAN

AC Q02653; MEDLINE: 90291512.

DR 01-JUL-1993 (Rel. 26, Created)

DR 01-JUL-1993 (Rel. 26, Last sequence update)

DR 15-JUL-1993 (Rel. 38, Last annotation update)

DE ATP SYNTHASE PROTEIN 8 (EC 3.6.1.34).

GN ATP8.

OS Podospora anserina.

OC Mitochondrion.

OC Eukaryota; Fungi; Ascomycota; Euascomycetes; Pyrenomycetes; Sordariales; Sordariaceae; Podospora.

CC [1]

CC SEQUENCE FROM N.A.

CC STRAIN-S:

CC MEDLINE: 90291512.

CC RA CUMMINGS D. J., MCNALLY K. L., DOMENICO J. M., MATSURA E. T.;

CC "The complete DNA sequence of the mitochondrial genome of Podospora anserina.";

CC Curr. Genet. 17:375-402(1990).

CC -1- FUNCTION: THIS IS ONE OF THE CHAINS OF THE NONENZYMATIC COMPONENT (CF(0) SUBUNIT) OF THE MITOCHONDRIAL ATPASE COMPLEX.

CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.

CC -1- SIMILARITY: BELONGS TO THE ATPASE PROTEIN 8 FAMILY.

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL: X97707; CA66287.1; -

CC DR PFAM: PF00895; ATP-synt_8.1.

CC KW Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane.

CC FT TRANSMEM 8 32 POTENTIAL.

CC SEQUENCE 50 AA; 5944 MW; D3E7DF1F CRC32;

ATP8_PODAN Length: 50 February 14, 2000 08:02 Type: P Check: 8763

1 MPQLVPEYFV NEITFEFIL AIVYLISKY IIPREVLFL SRPISRLIG

11AA_SEQUENCE 1.0 STANDARD; PRT: 68 AA.

ATP8_PONPA

AC P92694; MEDLINE: 97032590.

DR 01-NOV-1997 (Rel. 35, Created)

DR 01-NOV-1997 (Rel. 35, Last sequence update)

DR 15-JUN-1999 (Rel. 38, Last annotation update)

DE ATP SYNTHASE PROTEIN 8 (EC 3.6.1.34) (A65).

GN ATP8 OR ATP8.

CC Pongo pygmaeus abelii (Sumatran orangutan).

CC Mitochondrion.

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eukarya; Primates; Catarrhini; Homnidae; Pongo.

CC [1]

CC SEQUENCE FROM N.A.

CC STRAIN-YN93-312.

CC MEDLINE: 97032590.

CC XU X., ARNASON U.;

CC "The mitochondrial DNA molecule of Sumatran orangutan and a molecular proposal for two (Bornean and Sumatran) species of orangutan.";

CC J. Mol. Evol. 43:431-437(1996).

CC -1- FUNCTION: THIS IS ONE OF THE CHAINS OF THE NONENZYMATIC COMPONENT (CF(0) SUBUNIT) OF THE MITOCHONDRIAL ATPASE COMPLEX.

CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.

CC -1- SIMILARITY: BELONGS TO THE ATPASE PROTEIN 8 FAMILY.

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL: X97707; CA66287.1; -

CC DR PFAM: PF00895; ATP-synt_8.1.

CC KW Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane.

CC FT TRANSMEM 8 32 POTENTIAL.

CC SEQUENCE 68 AA; 7860 MW; B5CEA057 CRC32;

ATP8_PONPA Length: 68 February 14, 2000 08:02 Type: P Check: 3336

```

1 MPOINTTWP TLITPTLAL FLITOLKLN SHLHPPTPK FTKRPHAKP
51 WELKTRKYS PHSLPPOX

ID ATP8_PONPP STANDARD; PRT; 68 AA.
AC P32896;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE ATP SYNTHASE PROTEIN 8 (EC 3.6.1.34) (A6L).
GN MTATP8 OR ATP8.
OS Pongo pygmaeus pygmaeus (Bornean orangutan).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Pongo.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ISOLATES ANNA AND DENNIS;
RX MEDLINE: 97032590.
RA XU X., ARNASON U.;
RT "The mitochondrial DNA molecule of Sumatran orangutan and a molecular
RT proposal for two (Bornean and Sumatran) species of orangutan.";
RL 1. Mol. Evol. 43:431-437(1996).
CC -1- FUNCTION: THIS IS ONE OF THE CHAINS OF THE NONENZYMATIC COMPONENT
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC -1- SIMILARITY: BELONGS TO THE ATPASE PROTEIN 8 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X97715; CA66301.1; -
DR EMBL: X97711; CA66297.1; -
DR PFAM: PF00895; ATP-synt_8; 1.
KW Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane.
FT TRANSMEM 8 24 POTENTIAL.
SQ SEQUENCE 68 AA: 7910 MW; 8AAC1F6C CRC32;

ATP8_PONPP Length: 68 February 14, 2000 08:02 Type: P Check: 3460

1 MPOINTTWP TLITPTLAL FLITOLKLN SHLHPPTPK FTKRPHAKP
51 WELKTRKYS PHSLPPOX

ID ATP8_PONPP STANDARD; PRT; 68 AA.
AC P35584;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE ATP SYNTHASE PROTEIN 8 (EC 3.6.1.34) (A6L).
GN MTATP8 OR ATP8.
OS Pongo pygmaeus (Orangutan).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Pongo.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE: 95132634.
RA HORAI S., HAYASAKA K., KONDO R., TSUGANE K., TAKARATA N.;
RT "Recent African origin of modern humans revealed by complete sequences
RT of hominoid mitochondrial DNAs.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:532-536(1995).
CC -1- FUNCTION: THIS IS ONE OF THE CHAINS OF THE NONENZYMATIC COMPONENT
CC (CF(0) SUBUNIT) OF THE MITOCHONDRIAL ATPASE COMPLEX.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC -1- SIMILARITY: BELONGS TO THE ATPASE PROTEIN 8 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D38115; BAA07309.1; -
DR PFAM: PF00895; ATP-synt_8; 1.
KW Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane.
FT TRANSMEM 8 24 POTENTIAL.
SQ SEQUENCE 68 AA: 7820 MW; B5A77A6F CRC32;

ATP8_PONPP Length: 68 February 14, 2000 08:02 Type: P Check: 3550

1 MPOINTTWP TLITPTLAL FLITOLKLN SHLHPPTPK FTKRPHAKP
51 WELKTRKYS PHSLPPOX

ID ATPD_HUMAN STANDARD; PRT; 168 AA.
AC P30049;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DE ATP SYNTHASE DELTA CHAIN, MITOCHONDRIAL PRECURSOR (EC 3.6.1.34).
GN ATP5D.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92182007.
RA JORDAN E.M., BREEN G.A.M.;
RT "Molecular cloning of an import precursor of the delta-subunit of the
RT human mitochondrial ATP synthase complex.";
RL Biochim. Biophys. Acta 1130:123-126(1992).
RN [2]
RP SEQUENCE FROM N.A.
RA LAMBRDIN J.E., MCCREARY P.M., SKOWRONSKI E., ADAMSON A.W.,
RA BURKHART-SCHULTZ K., GORDON L., KYLE A., RAMIREZ M., STIMAGEN S.,
RA PHAN H., VELASCO N., GARNES J., DANGANAN L., POUNDSTONE P.,
RA CHRISTENSEN M., GEORGESCU A., AVILA J., LIU S., ATRIX C., ANDREISE T.,
RA TRANKHEIM M., AMICO-KELLER G., COFFIELD J., DUARTE S., LUCAS S.,
RA BRUCE R., THOMAS P., O'VAN G., KRONMILLER B., ARELLANO A.,
RA MONTGOMERY M., OW D., NOLAN M., TRONG S., KOBAYASHI A., OLSEN A.O.,
RA CARRANO A.V.;
RT Submitted (FEB-1998) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE OF 23-38.
RC TISSUE-LIVER;
RX MEDLINE: 93162045.
RA HOCHSTRASSER D.F., FRUTIGER S., PROUET N., BAISCH A., RAVIER F.,
RA PASQUILLI C., SANCHEZ J.-C., TISSOT J.-D., BELLOUVIST B., VARGAS R.,
RA APPEL R.D., HUGHES G.J.;
RT "Human liver protein map: a reference database established by
RT microsequencing and gel comparison.";
RL Electrophoresis 13:192-1001(1992).
CC -1- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC GRADIENT ACROSS THE MEMBRANE.
CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC SEEMS TO HAVE NINE SUBUNITS: A, B, C, D, E, F, G, F6 AND 8 (OR
CC A6L).
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC -1- SIMILARITY: BELONGS TO THE ATPASE SUBUNIT CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X63422; CAA45016.1; -
DR EMBL: X63423; CAA45017.1; -
DR EMBL: AC004221; AAC04304.1; -
DR PIR: S22348; S22348.
DR SWISS-ADPAGE: P30049; HUMAN.
DR MIM: 603150; -
DR PFAM: PF00401; ATP-synt-DE; 1.
KW ATP synthetase; CF(1); Hydrogen ion transport; Hydrolase;
KW Mitochondrion; Transit peptide.
FT TRANSLIT 1 22 MITOCHONDRION.
FT CHAIN 23 168 ATP SYNTHASE DELTA CHAIN.
SQ SEQUENCE 168 AA; 17490 MW; 0182AE71 CRC32;
ATPD_HUMAN Length: 168 February 14, 2000 08:02 Type: P Check: 1686 ..
1 MLEPALLRRP GIGRLVRRAR AYAEAAAP AASGPNOMSF TFASPTGVFF
51 NGANVQVAV PLITGAFGIL AAHVPLQVL RGLVVAE DGTSTKYFVS
101 SGAIANADS SVQLAEAV TLDMLDGA KANLEKAOE LVGTADATR
151 AEQIRIEAN EALYKALE
11AA_SEQUENCE 1.0 STANDARD: PRT: 127 AA.
ID ATPD_BACP3
AC P09354;
DT 01-MAR-1989 (Rel. 10; Created)
DT 01-MAR-1989 (Rel. 10; Last sequence update)
DT 01-NOV-1997 (Rel. 35; Last annotation update)
DE ATP SYNTHASE PROTEIN I.
OS Bacillus p33 (Thermophilic bacterium PS-3).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 86163679.
RA OHTA S., YOHDA M., ISHIZUKA M., HIRATA H., HANAMOTO T.,
RA OTAMURA-HANAMOTO Y., MATSUDA K., KAGAWA Y.,
RT "Sequence and over-expression of subunits of adenosine triphosphate
RT synthase in thermophilic bacterium PS-3".
RL Biochim. Biophys. Acta 933:141-155(1988).
CC -1- FUNCTION: A POSSIBLE FUNCTION FOR THIS PROTEIN IS TO GUIDE THE
CC ASSEMBLY OF THE MEMBRANE SECTOR OF THE ATPASE ENZYME COMPLEX.
CC -1- SIMILARITY: TO THE CORRESPONDING SUBUNIT IN OTHER BACTERIA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X07804; CAA30647.1; -
DR PIR: S01397; S01397.
KW Hydrogen ion transport; CF(0); Transmembrane.
SQ SEQUENCE 127 AA; 14595 MW; 5999D881 CRC32;
ATPD_BACP3 Length: 127 February 14, 2000 08:02 Type: P Check: 4392 ..
1 MGNLQAMFWR QVRILYLILA IYTLGFGTPT YKTVFSLIL GTSISLMMW
51 NLTKIKKFG QAVAAKKVR TLGLSLRAL AALAAYIVL YPOYHIYPT
101 VLGIMTSYIV IIDFFHKW KNDKQLA
11AA_SEQUENCE 1.0 STANDARD: PRT: 398 AA.
ID BMDA_XENLA

AC P25703;
DT 01-MAY-1992 (Rel. 22; Created)
DT 01-MAY-1992 (Rel. 22; Last sequence update)
DT 01-JUL-1993 (Rel. 26; Last annotation update)
DE BONE MORPHOGENETIC PROTEIN 2-1 PRECURSOR (BMP-2-1).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopodinae;
OC Xenopus.
CC [1]
CC SEQUENCE FROM N.A.
CC MEDLINE: 91274367.
CC PLESSEY S., KOESTER M., KNOECHER W.;
CC "cDNA sequence of Xenopus laevis bone morphogenetic protein 2
CC (BMP-2)".
CC Biochim. Biophys. Acta 1089:280-282(1991).
CC [2]
CC SEQUENCE FROM N.A.
CC MEDLINE: 92378616.
CC NISHIMATSU S., SUZUKI A., SHODA A., MURAKAMI K., UENO N.;
CC "Genes for bone morphogenetic proteins are differentially transcribed
CC in early amphibian embryos".
CC Biochem. Biophys. Res. Commun. 186:1487-1495(1992).
CC -1- FUNCTION: INDUCES CARTILAGE AND BONE FORMATION.
CC -1- SUBUNIT: HOMODIMER, DISULFIDE-LINKED.
CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X55031; CAA38850.1; -
DR EMBL: X63424; CAA45018.1; -
DR PIR: S16244; S16244.
DR PIR: JH0687; JH0687.
DR HSSP: P18075; BMP.
DR PROSITE: PS00250; TGF-BETA: 1.
DR PFAM: PF00019; TGF-beta: 1.
KW Signal; Growth factor; Cytokine; Bone; Cartilage; Glycoprotein.
RN SIGNAL 1 284
FT CHAIN 1 284
FT PROPEP 285 398
FT DISULFID 298 398
FT DISULFID 327 395
FT DISULFID 331 397
FT DISULFID 362 392
FT CARBOHYD 137 137
FT CARBOHYD 202 202
FT CARBOHYD 340 340
FT CONFLICT 7 7
FT CONFLICT 16 16
FT CONFLICT 233 233
SQ SEQUENCE 398 AA; 45575 MW; 6143F996 CRC32;
BMDA_XENLA Length: 398 February 14, 2000 08:02 Type: P Check: 9714 ..
1 WAGIHSLL LLFYQVLLSG CTGLPEEGK RRYTESGRSS POOSQRLVNO
51 FELRLSMFG LKRRTPGKN WVIPPYMLD YHLILAOLAA DEGTSMADFQ
101 MERAASRANT VRSFHEESM ELIPESEKRT IQRFFNLSS IPNEELVISA
151 ELRIPEQVO EPESDSSKL HRINIVDIK PAAASGQV VLRLDRLVH
201 HNSKKESED VTPALARMIA HKOPNGEYV EVNHLNDKN VPKKVRISR
251 SLTPDKNWP QIRPLVTF S HDGKGHALK RQKQARHKO RRLKSSCSR

```

301 HPLVDFSDV GWNWIVAPP GYHAFYCHGE CPEPLADHLN STNHAIVQTL
351 VNSVNTNIPK ACQVPELSA ISMLYLDENE KYVLKNYODM VVEGCGR

!!AA_SEQUENCE 1.0
BMBP_XENLA STANDARD: PRT: 398 AA.
ID BMBP_XENLA
AC P30884;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DE BONE MORPHOGENETIC PROTEIN 2-II PRECURSOR (BMP-2-II).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
  Batrachia; Anura; Mesobatrachia; Pipidae; Xenopodinae;
  OC Xenopus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92378616.
RA NISHIMATSU S., SUZUKI A., SHODA A., MURAKAMI K., UENO N.;
RT "Genes for bone morphogenetic proteins are differentially transcribed
  in early amphibian embryos."
RL Biochem Biophys Res Commun. 186:1487-1495(1992).
CC -1- FUNCTION: INDUCES CARTILAGE AND BONE FORMATION.
CC -1- SUBUNIT: HOMODIMER. DISULFIDE-LINKED.
CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X63425; CAA45019.1; -
DR F01888; JH0688.
DR HSSP: P18075; IBM.
DR PROSITE: PS00250; TGF_BETA: 1.
DR PFAM: PF00019; TGF-beta: 1.
KW Signal; Growth factor; Cytokine; Bone; Cartilage; Glycoprotein.
FT SIGNAL 1 284 ?
FT PROPEP 1 284 ?
FT CHAIN 285 398 ?
FT DISULFID 298 363 ?
FT DISULFID 327 395 ?
FT DISULFID 331 397 ?
FT DISULFID 362 362 ?
FT CARBOHYD 137 137 ?
FT CARBOHYD 202 202 ?
FT CARBOHYD 237 237 ?
FT CARBOHYD 340 340 ?
SQ SEQUENCE 398 AA: 45616 MW; E7DIDBA CRC32;

BMBP_XENLA Length: 398 February 14, 2000 08:02 Type: P Check: 1620

```

```

ID BPHR_PSEPS STANDARD: PRT: 275 AA.
AC P08654;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-NOV-1992 (Rel. 20, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE BI-PHENYL-2,3-DIHYDRO-2,3-DIOL DEHYDROGENASE (EC 1.3.1.-) (BI-PHENYL-
  CIS-DIOL DEHYDROGENASE) (2,3-DIHYDROXY-4-PHENYLHEXA-4,6-DIENE
  DEHYDROGENASE).
DE BPHR.
OS Pseudomonas pseudoalcaligenes.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
  OC Pseudomonas.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-KF707;
RX MEDLINE: 87083404.
RA FURUKAWA K., ARIMURA N., MIYAZAKI T.;
RT "Nucleotide sequence of the 2,3-dihydroxydiphenyl dioxygenase gene of
  Pseudomonas pseudoalcaligenes."
RL J. Bacteriol. 169:427-429(1987).
CC -1- PATHWAY: DEGRADATION OF BI-PHENYLS AND POLYCHLOROBIPHENYLS (PCB) TO
  BENZOIC ACID AND CHLOROBENZOIC ACIDS.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
  FAMILY (SDR).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M15333; AAA25752.1; ALT-INIT.
DR HSSP: P4722; 1BDB.
DR PROSITE: PS00061; ADH_SHORT: 1.
DR PFAM: PF00106; adh_short: 1.
KW Aromatic hydrocarbons catabolism; Oxidoreductase; NAD.
FT ACT_SITE 153 153 NAD (BY SIMILARITY).
FT SITE 153 153 ?
SQ SEQUENCE 275 AA: 28726 MW; EA999DF7 CRC32;

BPHR_PSEPS Length: 275 February 14, 2000 08:02 Type: P Check: 1527

```

```

1 MAGHSLL LQFOILISG CTGLVPEEGK RRYSESTSS PQOSQOVLDO
51 FERRLLMNG LKRRPTPEKN VVIPPWLDI YHLHSAOLAD DOGSSEVDYH
101 MERAASRANT VRSPHHEESM EEPESGERT IORFFNLSS IPDELYTSS
151 ELRIFREOVQ EPEFTDGSKL HRINIYDIK PAAASRGPV VRLDPLIH
201 HNESWESFD VTPAIRWIA HKOPNHGFV EYTHLDNDTN VPKRWISR
251 SLTLDKGHP RIRPLVTF S HGKGHALK ROKROARKQ RKLKSSCR
301 HPLVDFSDV GWNWIVAPP GYHAFYCHGE CPEPLADHLN STNHAIVQTL
351 VNSVNTNIPK ACQVPELSA ISMLYLDENE KYVLKNYODM VVEGCGR

!!AA_SEQUENCE 1.0

```

```

!!AA_SEQUENCE 1.0
ID BR3A_HUMAN STANDARD: PRT: 423 AA.
AC G01851; Q15318; Q14988.
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE BRAIN-SPECIFIC HOMEOBOX/POU DOMAIN PROTEIN 3A (BRN-3A) (OCT-11)
  (HOMEOBOX/POU DOMAIN PROTEIN RDC-1).
DE FOD4F1 OR BRN3A OR RDC1.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX Xiang M., Zhou L.-J., Macke J.P., Eddy R.L., Shown T.B., Nathans J.;
  RA Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.

```

[2]
 RN SEQUENCE FROM N.A.
 RP MEDLINE: 94052142.
 RA BHARGAVA A.K., LI Z., WEISSMAN S.M.;
 RT "Differential expression of four members of the Pou family of
 RT proteins in activated and phorbol 12-myristate 13-acetate-treated
 RT Burkitt T cells";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:10260-10264(1993).
 RN [3]
 RP SEQUENCE OF 85-423 FROM N.A.
 RC TISSUE-PLACENTA;
 RX MEDLINE: 93027214.
 RA COLLIUM R.G., FISHER P.E., DATTA M., MELLIS S., THIELE C.,
 RA HUEBER K., CROCE C.M., ISRAEL M.A., THEIL T., MOROY T., DEPINHO R.,
 RA ALT F.W.;
 RT "A novel Pou homeodomain gene specifically expressed in cells of the
 RT developing mammalian nervous system";
 RL Nucleic Acids Res. 20:4919-4925(1992).
 CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR WHICH MAY PLAY A ROLE IN
 CC THE REGULATION OF SPECIFIC GENE EXPRESSION WITHIN A SUBSET OF
 CC NEURONAL LINEAGES. MAY PLAY A ROLE IN DETERMINING OR MAINTAINING
 CC THE IDENTITIES OF A SMALL SUBSET OF VISUAL SYSTEM NEURONS.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- TISSUE SPECIFICITY: BRAIN. SEEMS TO BE SPECIFIC TO THE RETINA.
 CC PRESENT IN THE DEVELOPING BRAIN, SPINAL CORD AND EYE.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSION PEAKS EARLY IN EMBRYOGENESIS
 CC (DAY 13.5) AND IS UNDETECTABLE 14 DAYS AFTER BIRTH.
 CC -1- SIMILARITY: STRONG TO OTHER "POU" TRANSCRIPTION FACTORS. BELONGS
 CC TO CLASS-4 POU.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: 010063; AAA57161.1; -
 DR EMBL: 010062; AAA57161.1; JOINED.
 DR EMBL: L20433; AAA65605.1; -
 DR EMBL: X64624; CAA45907.1; -
 DR HSSP: P10037; 1A07.
 DR MIM: 601632; -
 DR PROSITE: PS00027; HOMEBOX_1; 1.
 DR PROSITE: PS00035; POU_1; 1.
 DR PROSITE: PS00465; POU_2; 1.
 DR PROSITE: PS00071; HOMEBOX_2; 1.
 DR PFM: PF00046; homeobox; 1.
 DR PFM: PF00157; pou; 1.
 DR KW Transcription regulation; Nuclear protein; DNA-binding; Homeobox;
 KM Developmental protein;
 FT DOMAIN 57
 FT DOMAIN 108
 FT DOMAIN 130
 FT DOMAIN 207
 FT DOMAIN 267
 FT DOMAIN 341
 FT DNA_BIND 359
 FT CONFLICT 91
 FT CONFLICT 92
 FT CONFLICT 99
 FT CONFLICT 133
 FT CONFLICT 135
 FT CONFLICT 137
 FT CONFLICT 139
 FT CONFLICT 149
 FT CONFLICT 155
 FT CONFLICT 174
 FT CONFLICT 189
 FT CONFLICT 240
 FT CONFLICT 252
 FT CONFLICT 256
 FT CONFLICT 276
 FT CONFLICT 343
 FT SEQUENCE 423 AA; 42939 MW; 0DC8F2FD CRC32;

BRAL_HUMAN Length: 423 February 14, 2000 08:02 Type: P Check: 2610 ..
 1 MMSMSKOPH FAMHPTLPEH KYPSPHSSE AIRRACLPTE PLOSMIFASL
 51 DELLARAELA LAAVDIAVSG GKSHPEKPD TYHTMSVPC TSTSTVPLAH
 101 HHHHHHHQA LEPGDLDDH SSPSLAMAG AGGAGGAGAA AGGGAHDGP
 151 GGGGGPGGG GPGGGPGGG GGGGGPGGG GPGGGGAGG AHPHHHMSL
 201 GHLSPAPAA AMMPGGLPH POLVAAAHH GAAAAAAGAA AGQVAAAAG
 251 AAAGAAAGLA SIDSDIDR ELEAFERK QRRKLGTV AVGSAALNL
 301 KIPGVSLSQ STICREBSLT LSHNMILK PLOAMLEEA EAQERKNNK
 351 PELFGGERK RKRSTIAPE KRLEAFYAV QPRPSSEKIA AIAERLCLK
 401 NVVRVFCNQ ROKOKRKES ATY
 !!AA_SEQUENCE 1.0
 ID BRAL_BRAFL STANDARD; PRT; 448 AA.
 AC 017134;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE BRACHYURY PROTEIN HOMOLOG 1 (AMBRA-1).
 GN BRA-1.
 OS Branchiostoma floridae (Florida lancelet) (Amphioxus).
 OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
 OC Branchiostoma.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-LARVA
 RX MEDLINE: 96125169.
 RA HOLLAND P.W.H., KOSCHORZ B., HOLLAND L.Z., HERRMANN B.G.;
 RT Conservation of Brachyury (T) genes in amphioxus and vertebrates:
 RT developmental and evolutionary implications";
 RL Development 121:4283-4291(1995).
 CC -1- FUNCTION: INVOLVED IN THE TRANSCRIPTIONAL REGULATION OF GENES
 CC REQUIRED FOR MESODERM FORMATION AND DIFFERENTIATION
 CC (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- DEVELOPMENTAL STAGE: FIRST DETECTED IN THE 10 H GASTRULA AROUND
 CC THE BLASTOPORE. IN LATER STAGES OF GASTRULATION AND EARLY
 CC NEURALIZATION, EXPRESSED IN POSTERIOR MESODERM AND THE NOTOCHORD
 CC WHERE EXPRESSION CONTINUES AT LEAST UNTIL SWIMMING LARVAL STAGE.
 CC -1- SIMILARITY: CONTAINS A T-BOX DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X91503; CAA62959.1; -
 DR HSSP: F24781; 1XBR.
 DR PROSITE: PS01264; TBOX_2; 1.
 DR PROSITE: PS01283; TBOX_1; 1.
 DR PFM: PF00907; T-box; 1.
 DR KW Developmental protein; Transcription regulation; DNA-binding;
 KM Nuclear protein; Multigene family.
 FT DNA_BIND 54
 FT SEQUENCE 448 AA; 49253 MW; 5948F2D3 CRC32;
 BRAL_BRAFL Length: 448 February 14, 2000 08:02 Type: P Check: 7365 ..
 1 MSAEMKOP TAASPDQFSV SHLSAVESE ISASEKGP TENDLITILE
 51 EPLMDKFA LTNEMITKN GRMPEVLKV NVSGIDBNAM YSLIDFTAA

101 DNRKMYNG EWPGRKPEP SVPSCVIHP DSPNFGAHMM KSPVSFSYVK
 151 LTNKLNKGGG IMLNSLHKYE PRHILIKYGG PGNQMWSTH TPEETOFIAY
 201 TAYONEIITA LKIKYNPAK AFIDAKERSD GKDGHEDDQ QPQYSLQGGN
 251 FLPGTGPICP PPNHQFAPS LGPISGCDR YSTLANHSA PYHPYORSS
 301 PPTNYGHDTA ASLPMPHTND NWSGLPVSTH NMLSMANPH TTSTTAQYF
 351 NLMVSNNML TPTTHAOTHM SGTMGGLPH QFLRTAPAP YHSIPICTYF
 401 TTASSPXYH DSHEVSTDS GYGHSTTPA POTRITSNM SPMTPEYL

11AA_SEQUENCE 1.0
 ID BRA2_BRAFL STANDARD: PRT: 440 AA.
 AC P80492:
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE BRACHYURY PROTEIN HOMOLOG 2 (AMBR-2).
 GN BRA-2.
 OS Branchiostoma floridae (Florida lancelet) (Amphioxus).
 OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
 OC Branchiostoma.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 96125169.
 RA HOLLAND P.W.H., KOSCHORZ B., HOLLAND L.Z., HERRMANN B.G.:
 RT "Conservation of Brachyury (T) genes in amphioxus and vertebrates:
 RL developmental and evolutionary implications."
 CC -1- FUNCTION: INVOLVED IN THE TRANSCRIPTIONAL REGULATION OF GENES
 CC REQUIRED FOR MESODERM FORMATION AND DIFFERENTIATION
 CC (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -1- TISSUE SPECIFICITY: MESODERM AND NOTOCHORD.
 CC -1- SIMILARITY: CONTAINS A T-BOX DOMAIN.
 DR HSSP: P24781; 1XBR.
 DR PROSITE: PS01264; TBOX_2; 1.
 DR PPM: PF00907; T-box; 1.
 DR Developmental protein; Transcription regulation; DNA-binding;
 KM Nuclear protein; Multigene family.
 FT DNA BIND 44
 FT SEQUENCE 440 AA; 48469 MW; 4828EB28 CRC32;

BRA2_BRAFL Length: 440 February 14, 2000 08:02 Type: P Check: 391

1 MKOTPOQFSV SHLSAASE ISAGSEKGP TERDLKVLG EKPLMEKFS
 51 LTNEMIVTYS GRMFPVLKV NVGLDPMAM YSLDLFTPA DNRKMYNG
 101 EWPGRKPEP SVPSCVIHP DSPNFGAHMM KSPVSFSYVK LTNKLNKGGG
 151 QIMLSLHKY EPRHIVKVG GPDNORTLS HTFAETOFIA VTAYQNEELT
 201 ALKIKHNPPA KAFIDAKERN DTKSGHDLT DOOPFOSLG GWELPGTPI
 251 CPPNPHOFA PSILGPHSGC DRYSTLRNH SAPYHPYOR SSPPTNYGHD
 301 TAASLPMPPT HDNWSGLPVSTH NMLSMANSA PHTTSTHAQ YPNLMVSNN
 351 NLPTTHAQT HMSGTGTGL PHQFLRTAP APYHSIPCT VPTASSPV
 401 YHSHSVST DSGYGHSTP PAROTRITS NMSPTNMSM

11AA_SEQUENCE 1.0
 ID BSR_BACCE STANDARD: PRT: 140 AA.
 AC P3367:
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)

01-NOV-1995 (Rel. 32, Last annotation update)
 DE BLASTICIDIN-S DEAMINASE (EC 3.5.4.23).
 GN BSR.
 OS Bacillus cereus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-46.
 RC STRAIN-K55-S1.
 RX MEDLINE: 92144117.
 RA KOBAYASHI K., KAKAKURA T., TANAKA T., YAMAGUCHI I., ENDO T.:
 RT "Nucleotide sequence of the bsr gene and N-terminal amino acid
 RT sequence of blastcidin S deaminase from blastcidin S resistant
 RT Escherichia coli TK121."

CC Agric. Biol. Chem. 55:3155-3157(1991).
 CC -1- FUNCTION: CATALYSES THE DEAMINATION OF THE CYTOSINE MOIETY OF THE
 CC -1- ANTIBIOTIC BLASTICIDIN S, CYTOMYCIN AND ACETYLBASTICIDIN S.
 CC -1- CATALYTIC ACTIVITY: BLASTICIDIN S + H(2)O =
 CC DEAMINOHYDROXYBLASTICIDIN S + NH(3).
 CC -1- COFACTOR: ZINC (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE CYTIDINE AND DEOXYCYTIDYLATE DEAMINASES
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch)
 CC -----

DR EMBL: S81409; AAC60404.1; -
 DR PIR: J50609; J50608
 DR PROSITE: PS00905; CYT_DCMP_DEAMINASES. 1.
 DR PPM: PF00383; dCMP_cyt_deam. 1.
 KM Hydrolyase; ZINC.
 FT METAL 59
 FT METAL 59
 FT METAL 100
 FT METAL 103
 FT METAL 103
 FT ZINC (BY SIMILARITY).
 FT ZINC (BY SIMILARITY).
 FT ZINC (BY SIMILARITY).
 SO SEQUENCE 140 AA; 15573 MW; 1208CACC CRC32;

BSR_BACCE Length: 140 February 14, 2000 08:02 Type: P Check: 8425

1 MKTNISQOD LELVEATEK ITMLEDNKH HVGCAIRTKT GEISAVHIE
 51 AYGRVTVCA EALVIGSAVS NGQKDEPTV AVRHPSDEV DRSIRVSPC
 101 GMRCLISDY APDCFLVLEM NGKLVKTTIE ELIPLKVTEN

11AA_SEQUENCE 1.0
 ID CADL_DICDI STANDARD: PRT: 213 AA.
 AC P34637:
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE CALCIUM-DEPENDENT CELL ADHESION MOLECULE-1 (DCCAD-1) (GP24).
 GN CADA.
 OS Dictyostelium discoideum (slime mold).
 OC Eukaryota; Dictyostelida; Dictyostelium.
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 43-62 AND 118-133.
 RX MEDLINE: 96279194.
 RA WONG E.F.S., SESAKI H., YANG C., SIU C.-H.:
 RT "Molecular cloning and characterization of Dccad-1, a Ca²⁺-dependent
 RT cell-cell adhesion molecule, in dictyostelium discoideum."
 RL J. Biol. Chem. 271:16399-16408(1996).
 CC -1- FUNCTION: INVOLVED IN CALCIUM-DEPENDENT CELL ADHESION.
 CC -1- SUBCELLULAR LOCATION: ASSOCIATED WITH THE ECTO-SURFACE OF THE
 CC PLASMA MEMBRANE. MAY BE TRANSPORTED TO THE PLASMA MEMBRANE VIA
 CC CONTRACTILE VACUOLAR AND ITS CELL SURFACE ASSOCIATION MAY BE
 CC MEDIATED BY AN INTEGRAL MEMBRANE PROTEIN.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED SOON AFTER THE INITIATION OF
 CC DEVELOPMENT.

CC -1- DOMAIN: CELL BINDING ACTIVITY IS DEPENDENT ON THE N-TERMINAL
 CC SEGMENT AND NOT ON THE PUTATIVE CA(2+)-BINDING SITE IN THE C-
 CC TERMINUS.
 CC -1- PTM: THE N-TERMINUS IS BLOCKED.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U49650; AAC47135.1; -
 CC DICTID: DD02037; CADA.
 CC Cell Adhesion; Calcium; Repeat.
 CC DOMAIN 1 184 4 X APPROXIMATE TANDEM REPEATS.
 CC REPEAT 1 97 1.
 CC REPEAT 49 97 2.
 CC REPEAT 98 146 3.
 CC REPEAT 147 194 4.
 CC REPEAT 213 AA; 23926 MW; DB554D11 CRC32:
 CC SEQUENCE
 CC CADI_DICDI Length: 213 February 14, 2000 08:02 Type: P Check: 3813 ..
 CC
 CC 1 MSVDANKKPF EFGKNCCTGES FEYKGETVR FNNCKRNDK FMSCLVGSNV
 CC 51 RCNIWEHNEI DPTPKRQFE LAQSTNNDL TSINGISKFO VLPQAFOMAV
 CC 101 DVKIVNKVNS TAGSEYMTIT PYQYDVACK DGDDEYOLPI PKLPPDSEI
 CC 151 VSHLTVAQTH TPYDYVNGS VYFKYSPTEG QYTVIKDET FPKNMTVTD
 CC 201 DNTSEIFNLN SEK
 CC
 CC 11AA_SEQUENCE 1.0 STANDARD; PRT; 294 AA.
 CC ID CANG_MOUSE
 CC AC P48070;
 CC DT 01-FEB-1996 (Rel. 33, Created)
 CC DT 01-FEB-1996 (Rel. 33, Last sequence update)
 CC DT 15-JUL-1998 (Rel. 36, Last annotation update)
 CC DE CALCIUM-SIGNAL MODULATING CYCLOPHILIN LIGAND (CAML).
 CC GN CALG OR CAML.
 CC OS Mus musculus (Mouse).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN-B6/CBA; TISSUE-THYMUS;
 CC RX MEDLINE; 96011657.
 CC RA KIM H.S., MORALES V.M., DASS C., ENCINAS J., TETTEL M.,
 CC RA BLUMBERG R.S.;
 CC RT "Cloning of the gene encoding the mouse homologue of the human
 CC calcium signal-modulating ligand."
 CC RL Gene 163:323-324(1995).
 CC CC -1- FUNCTION: LIKELY INVOLVED IN THE MOBILIZATION OF CALCIUM AS A
 CC RESULT OF THE TCR/CD3 COMPLEX INTERACTION. BINDS TO CYCLOPHILIN B.
 CC CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U21960; AAA87004.1; -
 CC MGD: MGI:104728; CAML.
 CC DR Transmembrane
 CC KM Transmembrane 188 208 POTENTIAL.
 CC FT TRANSMEM 239 255 POTENTIAL.
 CC SQ SEQUENCE 294 AA; 32514 MW; 6ED303C5 CRC32:

CANG_MOUSE Length: 294 February 14, 2000 08:02 Type: P Check: 283 ..
 CC
 CC 1 MEVPATATG GERPATPSGL SASORAEUR RRLKLNSEQ RINRWGFR
 CC 51 PGGSSEENG TSKRQPSDK LNSLSPSVS KRVLGDSVD GGGADLGCV
 CC 101 AFRGQQLQD KUDSPKAPK CSSKDAEELR QRTREDLTD PAORASHNL
 CC 151 EQLSREERA MKLRKQLSE KPSQEDGSTA EEPDEFIRFR LVGCALLALG
 CC 201 VRAVCKYIS IPAPITQL AYMGKYPF KGEKVKTV LTAALLSGI
 CC 251 PAEAINRMD TYSKMEYFT DLVYFTTFI FCHELDVG PEPV
 CC
 CC 11AA_SEQUENCE 1.0 STANDARD; PRT; 136 AA.
 CC ID CDD_BACSU
 CC AC P19079;
 CC DT 01-NOV-1990 (Rel. 16, Created)
 CC DT 01-NOV-1990 (Rel. 16, Last sequence update)
 CC DT 15-JUL-1999 (Rel. 38, Last annotation update)
 CC DE CYTIDINE DEAMINASE (EC 3.5.4.5) (CYTIDINE AMINOHYDROLASE) (CDA).
 CC GN CDD.
 CC OS Bacillus subtilis.
 CC OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC OC Bacillus/staphylococcus group; Bacillus.
 CC [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN-168;
 CC RX MEDLINE; 89313687.
 CC RA SONG B.-H., NEUHARD J.;
 CC RA "Chromosomal location, cloning and nucleotide sequence of the
 CC Bacillus subtilis cdd gene encoding cytidine/deoxycytidine
 CC deaminase."
 CC RL Mol. gen. genet. 216:462-468(1989).
 CC [2]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN-168 / JH62;
 CC RA KOBAYASHI Y., KIZONO M., MASUDA S., TAKEMARU K., HOSONO S.,
 CC RA SAITO T., TAKEUCHI M.;
 CC RL Submitted (MAR-1996) to the EMBL/Genbank/DBD databases.
 CC [3]
 CC RP SEQUENCE OF 1-10 FROM N.A.
 CC RC STRAIN-ED40;
 CC RA KIM K., HWANG S., SUH J., SONG B.-H., HONG S., KIM J.;
 CC RT "Nucleotide sequence upstream of the cdd locus in Bacillus subtilis."
 CC RL J. Microbiol. Biotechnol. 5:0-0(1995).
 CC CC -1- FUNCTION: THIS ENZYME SCAVENGE EXOGENOUS AND ENDOGENOUS CYTIDINE
 CC AND 2'-DEOXYCYTIDINE FOR UMP SYNTHESIS.
 CC CC -1- CATALYTIC ACTIVITY: CYTIDINE + H(2)O = URIDINE + NH(3).
 CC CC -1- COFACTOR: ZINC (BY SIMILARITY).
 CC CC -1- SUBUNIT: HOMOTETRAMER.
 CC CC -1- SIMILARITY: BELONGS TO THE CYTIDINE AND DEOXYCYTIDYLATE DEAMINASES
 CC FAMILY.
 CC CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U18532; AAB59993.1; -
 CC EMBL: X17432; NOT ANNOTATED_CDS.
 CC DR EMBL: D84432; BAA1248.1; -
 CC DR EMBL: U29117; AAB70045.1; -
 CC DR EMBL: K02114; AAB03947.1; -
 CC DR EMBL: Z89116; CAB14459.1; -
 CC DR EMBL: Z89117; CAB14472.1; -
 CC DR PIR: JEU022; JEU022.
 CC DR HSSP: P13652; IAF2.
 CC DR SUBTILIST; BG10477; CDD.

CDL_BACSU length: 136 February 14, 2000 08:02 Type: P Check: 9048 ..

1 MNRQELITPA LKARDMAVAP VSKFOYGAL LTKDSKYRG CNIENAYSM

51 CACAERTALE KAVEGDETF QMLAVADIP GPVSPGACR QVISELCTKD

101 VIVLTNLOG QIKETVEEL LPGAESSEDL HDERKL

11AA_SEQUENCE 1.0 STANDARD: PRT: 133 AA.

ID CDD_MYCPI

AC P47718:

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DE CYTIDINE DEAMINASE (EC 3.5.4.5) (CYTIDINE AMINOHYDROLASE) (CDA).

GN CDD.

OS Mycoplasma plium.

OC Bacteria; Firmicutes; Bacilli; Clostridium group; Molluscites;

OC Mycoplasmataceae; Mycoplasma.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BER:

RA MEDLINE: 9352438.

RA THAM T.N., PERRIS S., KOVACIC R., MONTAGNER L., BLANCHARD A.;

RA "Identification of Mycoplasma plium genes involved in the salvage

RA pathways for nucleosides".

RL J. Bacteriol. 175:5281-5285(1993).

CC -1- FUNCTION: THIS ENZYME CATALYZES EXOGENOUS AND ENDOGENOUS CYTIDINE

CC AND 2-DEOXYCYTIDINE FOR UMP SYNTHESIS (BY SIMILARITY).

CC -1- CATALYTIC ACTIVITY: CYTIDINE + H(2)O -> URIDINE + NH(3).

CC -1- COFACTOR: TIGHTLY BIND A SINGLE ATOM OF ZINC (BY SIMILARITY).

CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE CYTIDINE AND DEOXYCYTIDYLATE DEAMINASES

CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CDL_MYCPI length: 133 February 14, 2000 08:02 Type: P Check: 6718 ..

1 MKEKDIYFQK LNEILSNAYV PYSNFRVSCI LTTDGGWFG VNIENSAYSP

51 TICARSAYS SMITSGFQI FKVIYILDTI VKDIGTPOGV CROYLSFAK

101 PETPIITVNL KGEKFYTL EQLPFAFNKD ALK

11AA_SEQUENCE 1.0 STANDARD: PRT: 288 AA.

ID CDS_HAEIN

AC P44937: 032623: 032627:

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

15-JUL-1996 (Rel. 36, Last annotation update)

DE PHOSPHATIDATE CYTIDYLTRANSFERASE (EC 2.7.7.41) (CDP-DIGLYCERIDE

DE SYNTHASE) (CDP-DIGLYCERIDE PYROPHOSPHORYLASE) (CDP-DIACYLGlycerol

DE SYNTHASE) (CDS) (CTP:PHOSPHATIDATE CYTIDYLTRANSFERASE) (CDP-DAG

DE SYNTHASE).

GN CDS OR CDS OR H10919.

OS Haemophilus influenzae.

OC Bacteria; Proteobacteria; gamma subdivision: Pasteurellaceae;

OC Haemophilus.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-RD / KW20:

RX MEDLINE: 9530630.

RA FLEISCHMAN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,

RA KERAVAGE A.R., BUTT C.J., TOWB J.F., DOUGHERTY B.A., MERRICK J.M.,

RA MCKENNEY K., SUTTON G., FITZGUGH W., FIELDS C.A., GOCAYNE J.D.,

RA SCOTT J.D., SHIRLEY R., LIU L.-T., GLODER A., KELLEY J.M.,

RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,

RA UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDER D.M., BRANDON R.C.,

RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEORGHAGEN N.S.M.,

RA GENEH C.L., McDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,

RA VENTER J.C.;

RT "Whole-genome random sequencing and assembly of Haemophilus

RT influenzae Rd.";

RL Science 269:496-512(1995).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-33:

RA LOOSMORE S.M., YANG Y., COLEMAN D.C., SHORTREED J.M., ENGLAND D.M.,

RA KLEIN M.H.;

RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.

CC -1- CATALYTIC ACTIVITY: CTP + PHOSPHATIDATE -> PYROPHOSPHATE +

CC CDP-DIACYLGLYCEROL.

CC -1- PATHWAY: PHOSPHOLIPID BIOSYNTHESIS.

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE

CC (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE CDS FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: U32773: AAC22577.1; -

CC EMBL: U60831: AAB61967.1; -

CC EMBL: U60832: AAB61972.1; -

CC TIGR: H10919; -

DR PROSITE: PS01315; CDS: 1.

DR PFAM: Pf01148; Cytidylyltrans: 1.

DR TRANSFERASE; Nucleotidyltransferase; Phospholipid biosynthesis;

KW Transmembrane; inner membrane

KM TRANSMEM 10

FT TRANSMEM 10 POTENTIAL.

FT TRANSMEM 32 POTENTIAL.

FT TRANSMEM 89 POTENTIAL.

FT TRANSMEM 118 POTENTIAL.

FT TRANSMEM 152 POTENTIAL.

FT TRANSMEM 192 POTENTIAL.

FT TRANSMEM 223 POTENTIAL.

FT TRANSMEM 243 POTENTIAL.

FT TRANSMEM 16 V -> F (IN STRAIN EGAN).

FT TRANSMEM 59 T -> A (IN STRAIN 33).

FT TRANSMEM 81 R -> H (IN STRAIN EGAN).

FT VARIANT

FT SEQUENCE 288 AA: 32247 MW: 652F561 CRC32;

CDS_HAEIN length: 288 February 14, 2000 08:02 Type: P Check: 8973 ..

1 MKQRYLSAI VLIAAVICAL FLTPPYPAL ALGAVAIIGI WENQFARLK

51 OPLIRFVIT FLGVIFILML YTEGNYLDAG RVEBHDLL LINAVSWGL

101 ALLVYISPK SAKRWSKNPL LQLLFAPSTL IPFAGVILRL RLEHYTHDPY

151 HGLLELLYF ILWMADSGA YFSGRAGKR KLAHVSPCK SMEGIVGLLI
 201 TALVAFIETI HESNNLVGD RNINGFILLS VAVVAISVIG DLTESEKRE
 251 SGVNDSQLI PGHGVLDRI DSLTANVFF SYFFIFVL
 !!AA-SEQUENCE 1.0 STANDARD; PRT; 225 AA.
 ID CDSN_PIG
 DT 01-MAR-1992 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE CORNDESMOSIN (S PROTEIN) (FRAGMENT).
 OS CDSN.
 GN Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-PERIPHERAL BLOOD;
 RA VALIEN F.W., ROGEL-GAILLARD C., RENARD C., PONTAROTTI P.,
 RA TAZI-AHINI R., VAIMAN M., CHARON P.,
 RT Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- INTERCELLULAR LOCATION: SECRETED; FOUND IN CORNDESMOSOMES; THE
 CC INTERCELLULAR STRUCTURES THAT ARE INVOLVED IN DESQUAMATION (BY
 CC SIMILARITY).
 CC -!- TISSUE SPECIFICITY: EXCLUSIVELY EXPRESSED IN SKIN (BY SIMILARITY).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: Z97402; CAB10690.1;
 CC KW Glycoprotein.
 FT NON_TER 1
 FT DOMAIN 79 85 POLY-SER.
 FT DOMAIN 129 140 POLY-SER.
 FT DOMAIN 159 168 POLY-SER.
 FT DOMAIN 179 184 POLY-SER.
 FT NON_TER 225
 SEQUENCE 225 AA; 21644 MW; FE1295F0 CRC32;
 CDSN_PIG Length: 225 February 14, 2000 08:02 Type: P Check: 9217

1 GKPCEPITSV DKFGSYEVVG GSSDYLVPV MTYGGKIYP VGIYFKNPV
 51 KSGPAPSPA AGPISSEKY FSSNPPIPSH SSSSSNIYOS GASSAIVFOP
 101 VSGGAVOPCG VSSGSKGPC SLTSGSVHSS SSISSSGSS FHPGTVSOG
 151 PCSPGTGSF SSSSSKSGG KIILPCGSK SSSSGHPCIS VSSSTLSGGP
 201 DSGPDPSPA GAKPGSGSS GKIPC
 !!AA-SEQUENCE 1.0 STANDARD; PRT; 415 AA.
 ID CG21_SCHPO
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE G2/MITOTIC-SPECIFIC CYCLIN C1G1.
 GN C1G1.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 CC Schizosaccharomyces.
 CC [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 91300545.
 RA BUENO A., RICHARDSON H.E., REED S.I., RUSSELL P.;
 RT "A fission yeast B-type cyclin functioning early in the cell cycle."
 RL Cell 66:149-159(1991).
 CC -!- FUNCTION: REQUIRED FOR EFFICIENT PASSAGE OF THE G1/S TRANSITION.
 CC -!- SIMILARITY: BELONGS TO THE CYCLIN FAMILY. CYCLIN AB SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: M68881; AAA35288.1;
 CC PIR: A40316; A40316.
 DR HSSP: P20248; JUST.
 DR PROSITE: P500292; CYCLINS; 1.
 DR PFAM: PF00134; cyclin; 1.
 DR Cyclin; Cell cycle; Cell division; Mitosis.
 KW SEQUENCE 415 AA; 47827 MW; ED01A48C CRC32;
 CG21_SCHPO Length: 415 February 14, 2000 08:02 Type: P Check: 7358

1 MDVSTQTRRA TYFDENQLQ KHIIVYKKS HIKLNTGVA PKAVDNIQO
 51 QDEPTLIEGN NESSISSSTG DFEEDFAVQ DKEIERSI RSTPSIGD
 101 DLENRSGSD APEGILTHGK HRLPIEWT KEDIALAUSA ARLAANSP
 151 EDIETPSMW PYDEIRHY MSLERKLAP PPNVSVQOE IDWTRHMLV
 201 DWIVQVIRH RLPLEFLA VNLIDFLSI KVSLOKQVL VGSALLINC
 251 KYEIHPSI YNEAVVQGI FVDEILRAE RYMLLDFD ISWPGNSFL
 301 RIRIRASVD HDIRMLAKYL QEVILDEIF IGAHSIFLAA TAYLSMOML
 351 GHLMTQCHV YVSGYARQL KPCANIWEC LVDAPNHNA IYKYSENRM
 401 KVSAPAHNV YLSVI
 !!AA-SEQUENCE 1.0 STANDARD; PRT; 196 AA.
 ID CH19_DROGR
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE CHORION PROTEIN S19.
 GN CPL9 OR S19.
 OS Drosophila grimshawi (Fruit fly) (Idiomyla grimshawi).
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Hexapoda;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 88297142.
 RA MARTINEZ-CRUZADO J.C., SWIMMER C., FENERITAN M.G., KARATOS F.C.;
 RT "Evolution of the autosomal chorion locus in Drosophila. I. General
 RT organization of the locus and sequence comparisons of genes S15 and
 RT S19 in evolutionary distant species."
 RL Genetics 119:663-677(1988).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: X53422; CAA37506.1;
 CC

DR PIR: S06613; S06613.
 DR FLYBASE: FBgn0012320; Dgr1Cp19.
 KW Chorion
 SQ SEQUENCE 196 AA; 19826 MW; 135544F5 CRC32;
 CH19_DROCR Length: 196 February 14, 2000 08:02 Type: P Check: 4664 ..

1 MMTFATLAI ISACLAAGSC GGYSPGIGY GPINGLRRV SSICQSGD
 51 AAAAASAGS GDNGPEYIIA GGAPRYGSSQ NRPILLNSG YHGGLNDNG
 101 RIAQIVGGGR SLGHLGHL GGHLCGRIGC NYGGRIYRP FTVOGATIL
 151 LYPGQNSYR ISSPEYSKV ILPVRAAPV AKLYIPNNY GSQVGY

11AA_SEQUENCE 1.0 STANDARD; PRT; 213 AA.
 ID CHRR_RHOSH
 AC P0665;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE TRANSCRIPTIONAL ACTIVATOR CHRR.
 GN CHRR.
 OS Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
 CC Rhodobacter.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-2.4.1;
 RX MEDLINE: 95238260.
 RA SCHIRKE B.A., DONOHUE T.J.;
 RT "ChR positively regulates transcription of the Rhodobacter
 RT sphaeroides cytochrome c2 gene."
 RL J. Bacteriol. 177:1929-1937(1995).
 RN [2]
 RP REVISIONS.
 RA NEWMAN J., DONOHUE T.J.;
 RL Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: POSITIVELY REGULATES TRANSCRIPTION OF THE CYC2 GENE
 CC FOR CYTOCHROME C2. MAY PLAY A ROLE IN THE INACTIVATION OF THE POE
 CC GENE.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: U11283; AAB17905.1;
 CC KW Transcription regulation; Activator; DNA-binding.
 CC FT VARIANT 38 38 C -> R (IN CHRR MUTANT; LOSS OF
 CC FUNCTION).
 CC SQ SEQUENCE 213 AA; 22865 MW; 0AAE7676 CRC32;

CHRR_RHOSH Length: 213 February 14, 2000 08:02 Type: P Check: 8138 ..

1 MTRHHVSDA LLTAYAAAGTL SEAFSLVYAT HSLDCERA RAGALDAVAG
 51 SLMEETAPVA LSEGLASVY AQLDRQIOP APARRADPPA PAPIADYVGR
 101 RLEUDVRRTL GGGVROAILP TGGEAIARLL WIPGGQAVPD HGRGLELTL
 151 VIGAFPFDET DRFGAGDIEI ADQLEHTHPV AERGLDCLCL AATAPLARFN
 201 SFLPKLVOPF FRI

11AA_SEQUENCE 1.0 STANDARD; PRT; 351 AA.
 ID COAT_TCV
 AC P06663;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE COAT PROTEIN.
 OS Turnip crinkle virus (TCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tombusviridae;
 CC Carnovirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 87283926.
 RA STOCKLEY P.G., MORRIS T.J.;
 RT "Structure and assembly of turnip crinkle virus. IV. Analysis of the
 RT coat protein gene and implications of the subunit primary
 RT structure."
 RL J. Mol. Biol. 194:265-276(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 89243179.
 RA CARRINGTON J.C., HEATON L.A., ZUIDEMA D., HILLMAN B.I., MORRIS T.J.;
 RT "The genome structure of turnip crinkle virus."
 RL Virology 170:219-226(1989).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (3.2 ANGSTROMS).
 RX MEDLINE: 87112726.
 RA HOGUE J.M., MAEDA A., HARRISON S.C.;
 RT "Structure and assembly of turnip crinkle virus. I. X-ray
 RT crystallographic structure analysis at 3.2-A resolution."
 RL J. Mol. Biol. 191:625-638(1986).
 CC -1- SIMILARITY: COAT PROTEINS FROM CARNOVIRUSES (CARMV, MNSV, TCV),
 CC DIANTHOVIRUSES (RCMV), SOBEMOVIRUSES (MCMV, SBMV), TOMBSVIRUSES
 CC (AMCV, CMV, CYRSV, AND TBSV), AND TNV ARE RELATED IN THEIR S
 CC REGION.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: X05193; CA28823.1;
 CC DR EMBL: M22445; AAB96971.1;
 CC DR PIR: JAO111; VCVERC.
 CC DR PIR: S07285; S07285.
 CC DR HSP: P11785; 2TBV.
 CC DR PROSITE: P80053; ICOSAH_VIR_COAT_S; 1.
 CC DR PFAM: PF00729; Viral_coat; 1.
 CC KW Coat protein.
 CC FT DOMAIN 1 80 R DOMAIN, INTERACTION WITH RNA.
 CC FT DOMAIN 82 238 S DOMAIN, VIRION SHELL.
 CC FT DOMAIN 239 351 P DOMAIN, PROTECTING.
 CC FT CONFLICT 252 252 D -> Q (IN REF. 2).
 CC SQ SEQUENCE 351 AA; 38052 MW; 064C2069 CRC32;

COAT_TCV Length: 351 February 14, 2000 08:02 Type: P Check: 1818 ..

1 MENDPRVRF ASDGAOMAIK WOKKWSITL SROQTARAA MGKLSFPAQ
 51 PVQKTRLSA PVALAYREVS TOPRVSTAND GTRSGSELI TLIKNTDTE
 101 PRYTAVALNP SEPTFNOLI KEAAYEKYR FTSLRFYSP MSPSTTGKRY
 151 ALAFEDRAK PPRNDIASLY NIEGCVSSVP WTGELITVPT DSTDFEVADG
 201 ISDPKLVGK KLIMATYGOG ANDAQLGEV RVEYTVOLKN RIGSTDADI
 251 GDFAGYKGP RLVWSKITG TAGEMDCHF LGTGNPSITL FYKRAVSGSL
 301 ENNDASDFSV LGEAAGSVQ WAGVYAERG QGVKMTTEE QPRKLOAIR
 351 I

11AA_SEQUENCE 1.0 STANDARD; PRT; 237 AA.
 ID COAT_TOBSV

TOHO T., AIBA H., BABA T., FUJITA.., HAYASHI K., INADA T.,
RA ISONO K., KASAI H., KIMURA S., KITAKAWA M., KITAGAMA M.,
MAKINO K., MIKI T., MIZUBUCHI K., MORI H., MORI T., MOTOMURA K.,
RA NAKADE S., NAKAMURA Y. NASHIMOTO H., NISHIO Y., OSHIMA T.,
SAITO N., SAMPEL G., SEKI Y., SIYASUNARARA S., TAGAMI H.,
TAKEDA J., TAKEMOTO K., YADA C., YANAMOTO Y., HORICHI T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 40,1-50.0 min region on the linkage map."
DNA Res. 3:379-392(1996).
[4]
RN SEQUENCE OF 1-188 FROM N.A.
RC STRAIN-K12 / DHS-ALPHA;
RA COLLINS C.M., GUTMAN D.N., ISAKA J.;
RL submitted (XXX-11993) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CATALYZES THE SYNTHESIS OF 5'-6-DIMETHYLBENZIMIDAZOLE
CC (DMB), THE LOWER LIGAND OF B12 AND THE TRANSFER OF RIBOSE
CC PHOSPHATE FROM NICOTINATE MONONUCLEOTIDE TO DMB
CC -1- CATALYTIC ACTIVITY: BETA-NICOTINATE D-RIBONUCLEOTIDE +
CC DIMETHYLBENZIMIDAZOLE = NICOTINATE + N1-(5'-PHOSHO-ALPHA-D-
CC RIBOSYL)-5,6-DIMETHYLBENZIMIDAZOLE.
CC -1- PATHWAY: COBALAMIN BIOSYNTHESIS, PART II AND III (COBI1 / COBIIL).
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation. Its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U33333; AAA78908.1; -
CC DR EMBL: AF0000291; AAC75052.1; -
DR EMBL: D80837; CAB21728.1; -
DR EMBL: D90638; CAB21739.1; -
DR EMBL: L25054; AAA56876.1; -
CC EMBL: B51251; COBT.
KW COBALAMIN biosynthesis; Porphyrin biosynthesis; Transferase;
KM Glycosyltransferase
FT CONFLICT 271 A -> L (IN REF. 1).
FT CONFLICT 301 L -> V (IN REF. 1).
SQ PROCEDE 359 AA; 36987 MW; 86515DE7 CRC32;
CORR_ECOLI Length: 359 February 14, 2000 08:02 Type: P Check: 7029 ..

1 MOULADILNT IPATIDSTAMS RAORHIDGL KPVSLSCKLE VLAIOLAMP
51 GLNGIFHYVK KAVLMCADH GWMEBGVALS PKREYTAIQAE NMTRGTTCVC
101 VLAEOGANY HVIVDGIDA EPPIGLINKR VARGSGNIAS AMARRRAE
151 KLILDVICYT QELANKGVTL FVGELGMAN TTPAALIVST TIGRPDEVV
201 GIGANTETDK LANIKDIVVR ATTLPNPDO DGVDVLAKVG GPDLVIAGV
251 MLGAASGCLP VLLDGLFLSYA ALALACOMP AIRPYLIPI SH LSNEKGARIA
301 LSHGLEPFL NMEMRLGECS GAALAMPIIE AACATYNNMG ELAASNIVLP
351 GNTSDLSN

11AA-SEQUENCE 1.0
ID COR3 SCHCO STANDARD: PRF: 268 AA.
AC PIA058;
DA 01-JAN-1990 (Rel. 13, Created)
DI 01-FEB-1995 (Rel. 33, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
GN CITROCHROME C OXIDASE POLYPEPTIDE III (Ec 1.9.3.1).
DE COII.
OS Schizopphyllum commune (Bracket fungus).
OG Mitochondrion.
OC Eukaryota, fungi; Basidiomycota; Hymenomycetes; Stereales;
Schizophyllaceae; Schizopphyllum.

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 89106272.
RA PHELPS L.G., BURKE J.M., DULRICH R.C., NOVOTNY C.P.;
RT "Nucleotide base sequence of the mitochondrial coII gene of
RT Schizosaccharomyces pombe."
RT Curr. Genet. 14:401-403(1988).
CC -1- FUNCTION: SUBUNIT I, II, AND III FORM THE FUNCTIONAL CORE OF
CC THE ENZYME COMPLEX.
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O +
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 3 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M36270; CAB25813.1; -
DR PIR: S07557; S07557.
DR HSSP: P00415; 10CC.
DR PFAM: PF00510; COX3; 1.
DR Oxidoreductase; Mitochondrion; Transmembrane.
KW SEQUENCE 268 AA; 29922 MW; 25E315FE CRC32;
COX3_SCHCO Length: 268 February 14, 2000 08:02 Type: P Check: 7659 ..

1 MTSIKFYQSF SAHLVQHPW PLVFSFLFN LAIGTVLNM GYSHSTTFD
51 LGAVTVGSI LMTVDIVIE GSFLEDRTKQ VQGLIIGTI LFISEVAF
101 ISVFWAFHS ALSPAVELGS TWPVGGIPL DTFSLPLENT ILLSSGAFV
151 TYGHNAIFSG KRLLDSITGLF LTVALLAIFS YFOAFETIHA GFSMSDVFV
201 TYFFASTGTH GIHWMLGTLF LEVSELRQVN YQTKENHNG LETSILYWHF
251 VDLVWLFEL VYFWGGA

!!AA-SEQUENCE 1.0
ID COX3_SCHPO STANDARD: PRT: 164 AA.
AC P79010.
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-NOV-1997 (Rel. 35, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE IV PRECURSOR (EC 1.9.3.1).
GN COX4.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Aschiomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RA KAWAMUKAI M.;
RT Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O +
CC -1- SUBUNIT: COMPOSED OF AT LEAST 11 SUBUNITS.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE VB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB000399; BAA19097.1; -

```

```

DR PROSITE: PS00848; COX5B; 1.
DR PFAM: PF01215; COX5B; 1.
KW Oxidoreductase; Mitochondrion; Transit peptide; Zinc.
RT TRANSIT
RT CHAIN 1 33 MITOCHONDRION (BY SIMILARITY).
FT METRL 34 164 CYTOCHROME C OXIDASE POLYPEPTIDE IV.
FT METRL 120 120 ZINC (POTENTIAL).
FT METRL 143 143 ZINC (POTENTIAL).
FT METRL 146 146 ZINC (POTENTIAL).
SQ SEQUENCE 164 AA; 18161 MW; DA19B411 CRC32;
COX4_SCHPO Length: 164 February 14, 2000 08:02 Type: P Check: 7810 ..

```

```

1 MEMSMRLRV RORAAVSTV SLVRFVGS IRREQNVVK AAOELANAK
51 EPSDLIGPG RDEVEPTDLE QATGDEREL LSELGSRDAF DKRPDASRK
101 GLTIDPIWVT SLDPYRHIC TGSPGSHNL IMWTVYKDKL RCPGCGSVY
151 KLFKMGDPNA EHSK

!!AA-SEQUENCE 1.0
ID CPQ1_CANPG STANDARD: PRT: 114 AA.
AC P81580;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CUTICLE PROTEIN CP1158 (CPQ1158).
OS Cancer pagurus (Rock crab).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura; Cancridae;
OC Cancridae; Cancridae; Cancer.
RN [1]
RP SEQUENCE.
RC TISSUE-CUTICLE;
RA ANDERSEN S.O.;
RT "Exoskeletal proteins from the crab, Cancer pagurus.";
RT Comp. Biochem. Physiol. 0:0-0(1999).
CC -1- TISSUE SPECIFICITY: CALCIFIED SHELL.
CC Structural protein; Cuticle; Repeat.
FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
FT REPEAT 1 17 1.
FT REPEAT 26 43 2.
FT REPEAT 87 70 3.
FT REPEAT 95 112 4.
SQ SEQUENCE 114 AA; 11601 MW; FC0D861C CRC32;
CPQ1_CANPG Length: 114 February 14, 2000 08:02 Type: P Check: 1605 ..

1 QVYSGIVSP DGNNIQPTHD FAHSYLKGP SGIVTSQGN LQTLAQASL
51 QAAAPAPLP VSHYVVSQGS VVGPGIYSP SGNVQFSHER ADNVVLYVPS
101 GIVTKDGNL QLRA

!!AA-SEQUENCE 1.0
ID CPFR_PEA STANDARD: PRT: 402 AA.
AC P21727;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE TRIOSE PHOSPHATE/PHOSPHATE TRANSLCATOR, CHLOROPLAST PRECURSOR (CPPT)
DE Pisum sativum (Garden pea).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eustoids I; Fabales; Fabaceae; Papilionoideae;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 73-98; 249-260; 269-289 & 324-350.
RA STRAIN: CV. PROGRESS NO. 9; TISSUE-SEEDLING;
RX MEDLINE: 91035597.
RA SCHNEEL D.J., BLOBEL G., PAIN D.;
RT "The chloroplast import receptor is an integral membrane protein of

```

```

RC chloroplast envelope contact sites."
RL J. Cell Biol. 111:1825-1838(1990).
RP
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 73-82.
RC
RN [2]
RP
RC TISSUE-LEAF:
RA WILLEY D.L., FISCHER K., WACHTER E., LINK T.A., FLUEGE U.-I.;
RA Molecular cloning and structural analysis of the phosphate
RA translocator from pea chloroplasts and its comparison to the spinach
RA phosphate translocator";
RA Planta 183:451-461(1991).
RL
RN [3]
RP
RC SEQUENCE FROM N.A.
RA WEBER A.;
RA Submitted (SEP-1992) to the EMBL/GenBank/DBJ databases.
RL
CC -1- FUNCTION: MEDIATES THE EXPORT OF FIXED CARBONS FROM THE
CC CHLOROPLASTS INTO THE CYTOSOL IN THE FORM OF TRIOSE PHOSPHATES.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN, LOCATED IN
CC ZONES OF CONTACT BETWEEN THE INNER AND OUTER MEMBRANE OF THE
CC CHLOROPLAST.
CC -1- SIMILARITY: SPONG, TO OTHER PLANTS CPT. ALSO SIMILAR TO
CC YEAST PROTEIN SLX1.
CC -1- CAUTION: WAS ORIGINALLY THOUGHT TO FUNCTION AS A CHLOROPLAST
CC PROTEIN IMPORT RECEPTOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/)
CC or send an email to license@sib-sib.ch)
CC -----
CC EMBL: X54639; CAJ38451.1; -
CC DR EMBL: X68077; CAJ46210.1; -
CC KW Transmembrane; Chloroplast; Transit peptide; Transport.
CC Transmembrane: 1 72
CC TRANSIT 73 402
CC CHAIN 73 96
CC FT DOMAIN 97 117 CHLOROPLAST.
CC FT DOMAIN 118 129 TRIOSE PHOSPHATE/PHOSPHATE TRANSLOCATOR.
CC FT DOMAIN 130 150 INTERMEMBRANE SPACE (POTENTIAL).
CC FT DOMAIN 151 207 POTENTIAL.
CC FT DOMAIN 208 228 INTERMEMBRANE SPACE (POTENTIAL).
CC FT DOMAIN 229 272 POTENTIAL.
CC FT DOMAIN 273 292 POTENTIAL.
CC FT DOMAIN 293 370 INTERMEMBRANE SPACE (POTENTIAL).
CC FT DOMAIN 371 391 POTENTIAL.
CC FT DOMAIN 392 402 POTENTIAL.
CC FT VARIANT 227 227 I -> L.
CC FT SEQUENCE 402 AA: 43649 MW: 43649 MW; ACCIEF795 CRC32:
CC
CC CPTN_PEA Length: 402 February 14, 2000 08:02 Type: P Check: 5889
CC
CC 1 MESRVLSRAIT TLSSLPFLINK LHRPLPANA LPSTKSGSV SGGSLWNGR
CC 51 QLSPELCEPV LKKGASILRP CPATAGGNDG AGEKRAPVG FSRPILAIT
CC 101 GFEFTWYFL NWIINIINKK IYNYEPIYFE VSVIHLAVG VYCLVSWTVG
CC 151 LPRRAPIDGN LKLLIPVAV CHALGHVTSN VSFAVAVSF THTVAKLEPF
CC 201 FMAASGFIL GOSIPITLML SLAPVYIGVS MASITELSFN WLGFIAMIS
CC 251 NISTYISIV SKKAMTMDG TINIYAVISII ALIICIPPAL IIEGPIILKT
CC 301 GFNDALKVG LYKFSYDLFW VGMFHYLNG VAVNLEKVA PLTHAAGNVL
CC 351 KRVEVIFST IIFGNKISTQ TGIGTGIALA GVALYSFKA QIBEERQAK
CC
CC 401 AA

```

```

ID IAA_SEQUENCE 1.0 STANDARD; PRT; 331 AA.
AC CSKR_ECOLI
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE SCCR_OPERON REPRESSOR (CSC OPERON REGULATORY PROTEIN).
GN SCCR.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-EC3132;
BOCKMANN J.;
RL submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: REPRESSOR FOR THE CSC OPERON. BINDS D-FRUCTOSE AS
AN INDUCER (BY SIMILARITY)
-1- SIMILARITY: BELONGS TO THE LACI FAMILY OF TRANSCRIPTIONAL
REGULATORS.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collabora-
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstat-
CC CC the European Bioinformatics Institute. There are no restrictions on
CC CC use by non-profit institutions as long as the content is in no
CC CC modified and this statement is not removed.
CC CC entities requires a license agreement (See http://www.isb-sib.ch/anno
CC CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X81461; CMA57220.1; -.
DR HSSP; P03023; ILCD.
DR PROSITE; P500356; HTH_LACI_FAMILY; 1.
DR PFAM; PF00356; lacI; 1.
DR PFAM; PF00532; Peripla_bp_like; 1.
KW Transcription 4 DNA BIND 23 H-T-H MOTIF (POTENTIAL).
FT SEQUENCE 331 AA; 36470 MW; 3A39A271 CRC32;
SQ
CSKR_ECOLI Length: 331 February 14, 2000 08:02 Type: P Check: 8969

1 MASLKDVARL AGVSMKATYSR VMNNAESVSP ATPRVLQAI QNLNYPVPLS
51 AKKAQAQGRK PSTLNVLAAD TATTPRSVDI LLAIEOTASE FQMSNFLINI
101 FEEDDAARAA RQLLAIRPDG IITYTMGLRI ITLEPSLYGE NYLVANCVAD
151 DPALPSYIPD DYTAQYESIQ HULACYRQD LCFMIPESAL ANGYRQSGE
201 QAWKDAQGRDL AEVYQFHMAT GDHHTYDLAS LINAIFSGK PDDVDVLGN
251 DRAAFAYAYQ LLAAGYRIFQ DVAVMGEDNL VGVGHFLUPP LTITQLPHFI
301 IGRFALHIT EGRGGRVTR IPCPLLIGS T

1 IAA_SEQUENCE 1.0 STANDARD; PRT; 297 AA.
ID CTRA_BACFI
AC 004443;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE CYTOCHROME Aa3 CONTROLLING PROTEIN.
GN CTRA.
OS Bacillus firmus.
OC Bacteria; firmicutes; Bacillus/Clostridium group;
CC Bacillus/staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-OF4;
RC MEDLINE; 93107080
RA QDIRK P.G., HICKS D.B., KRUITICH T.A.;
RT "Cloning of the cta operon from alkaliphilic Bacillus firmus OF4 and
characterization of the pH-regulated cytochrome caa3 oxidase it
encodes.";

```


RT J. Biol. Chem. 268:678-685(1993).
 CC -1- FUNCTION: CTA IS REQUIRED FOR CYTOCHROME A3 BIOSYNTHESIS AND
 CC INDEPENDENTLY FOR SPOULATION.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M94110; AAA2362.1; -
 DR PIR: A45335; A45335.
 KW Sporulation; Membrane.
 SQ SEQUENCE 297 AA; 32326 MW; BA198CCI CRC32;

CTA_BACFI Length: 297 February 14, 2000 08:02 Type: P Check: 716 ..

1 MKRRLKITSV ITSGIVIAL LGALVTKTG SGEGCATWP LCFGEVPTN
 51 PAETIEYS HRIVSLAGA MIITAIAM KQKHMEKAL ALFAAVILI
 101 ISGILGAGA AVFGOSKAIL ALHFGISAMS LAAVLLITLI AFEDGREHTM
 151 APKVSRGKY YVEFVITTCY AVTSGAVYK HSEATLACAG FPLNGOITP
 201 GLXGVGNHY FHRVGTILL LFLTIMWT LSRVRYRYL TWAVLSFL
 251 VVGOFISIS IYTONALSV GILHALISI LFSALSYMT IITPSH

!!AA_SEQUENCE 1.0 STANDARD; PRT; 306 AA.
 ID CTA_BACFI
 AC P12946;
 DT 01-JUN-1990 (Rel. 13, Created)
 DT 15-JUL-1998 (Rel. 13, Last sequence update)
 DE CYTOCHROME A3 CONTROLLING PROTEIN.
 OS CTA.
 NC Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 89359135.
 RA MOELLER J.P., TABER H.W.;
 RT "Isolation and sequence of cta, a gene required for cytochrome a3
 RT biosynthesis and sporulation in Bacillus subtilis."
 RL J. Bacteriol. 171:4967-4978(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RA BERTERO M., PRESECAN E., GLASER P., RICHOU A., DANCHIN A.;
 RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CTA IS REQUIRED FOR CYTOCHROME A3 BIOSYNTHESIS AND
 CC INDEPENDENTLY FOR SPOULATION.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M23915; AAA50254.1; -
 DR EMBL: 298682; CAB1340.1; -
 DR EMBL: 299111; CAB13360.1; -
 DR PIR: A33960; A33960.
 DR SUBLIST: Bg10213; CTA.
 KW Sporulation; Membrane.

SQ SEQUENCE 306 AA; 34085 MW; 366EB369 CRC32;
 CTA_BACSV Length: 306 February 14, 2000 08:02 Type: P Check: 1522 ..

1 MNKALALGV LTFVMDIV LGALVTKTG SGQCGGNWP LCHGRFPPEL
 51 NPASIEWSH RFASGISIIL VSLAFWSMR KIPIFRETT FLAISIIFL
 101 FLAALGALA VFGSNALIM ALHFGISLIS FASVLLITLI IFADKSVRT
 151 LVKPLQIGK MEFMIGILI YSYIVYGA YRHTESSLA CPNPPLCSPL
 201 NNGLPLOFHE WYQGHRAA LLLFWIIVA AVAITSYKD QKQIFWGNIS
 251 CLIFTLQAL SGIMITYSEL ALGFALHSF FICLFGVLC YFLLIARFR
 301 YESROS

!!AA_SEQUENCE 1.0 STANDARD; PRT; 127 AA.
 ID CU04_BACR
 AC P80675;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE CURTICE PROTEIN 4 (BC-NC4).
 OS Blaberus craniifer.
 NC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
 CC Blaberoidea; Blaberidae; Blaberus.
 RN [1]
 RP SEQUENCE
 RC TISSUE-POLYRIC CAECA;
 RX MEDLINE: 97218697.
 RA JENSEN U.G., ROTHMANN A., SKOU L., ANDERSEN S.O., ROEPSTORFF P.,
 RA HOLTRUP P.;
 RT "Cuticular proteins from the giant cockroach, Blaberus craniifer."
 RL Insect Biochem. Mol. Biol. 27:109-120(1997).
 KW Structural protein; Cuticle; Repeat.
 FT MOD_RES 1 1
 FT REPEAT 31 39
 FT REPEAT 84 92
 FT SEQUENCE 127 AA; 13383 MW; 26A82D7C CRC32;
 CU04_BACR Length: 127 February 14, 2000 08:02 Type: P Check: 1926 ..

1 QAVLPSIYS ILSKSVRYQ ELEPEGANV PADTKAEIK EKEHEVITHG
 51 LPVPSAVQ PIVSVVNP GVIPAVVPD APAPADTKA EISDNVAVV
 101 HAISPLSVV SHVPSPLVH VQVYVY

!!AA_SEQUENCE 1.0 STANDARD; PRT; 92 AA.
 ID CY2_RHOE
 AC P00098;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE CYTOCHROME C2.
 OS Rhodocyclus tenuis (Rhodospirillum tenue).
 CC Bacteria; Proteobacteria; beta subdivision; Rhodocyclus group;
 CC Rhodocyclus.
 RN [1]
 RP SEQUENCE.
 RC STRAIN-3761;
 RX MEDLINE: 79199668.
 RA AMBLER R.P., MEYER T.E., KAMEN M.D.;
 RT "Anomolies in amino acid sequences of small cytochromes c and
 RT cytochromes c' from two species of purple photosynthetic bacteria."
 RL Nature 278:661-662(1979).
 CC -1- FUNCTION: CYTOCHROME C2 IS FOUND MAINLY IN PURPLE, NONSULFUR,
 CC PHOTOSYNTHETIC BACTERIA WHERE IT FUNCTIONS AS THE ELECTRON DONOR
 CC TO THE OXIDIZED BACTERIOCHLOROPHYLL IN THE PHOTOPHOSPHORYLATION
 CC PATHWAY. HOWEVER, IT MAY ALSO HAVE A ROLE IN THE RESPIRATORY CHAIN

AND IS FOUND IN SOME NONPHOTOSYNTHETIC BACTERIA.
 -1- SIMILARITY: THIS SEQUENCE IS MORE CLOSELY RELATED TO THE SEQUENCES OF CYTOCHROME C551 FROM PSEUDOMONAS AND AZOBACTER THAN TO THE SEQUENCES OF CYTOCHROME C2 FROM OTHER SPECIES OF RHODOSPIRILLUM.
 PIR: A00090; CCOR2T.
 DR HSP: P00099; 2PAC.
 DR PROSITE: P500190; CYTOCHROME_C; 1.
 DR PRAM: PF00034; cytochrome_c; 1.
 KW Electron transport; Photosynthesis; Heme.
 FT BINDING 12 12 HEME (COVALENT).
 FT BINDING 15 15 HEME (COVALENT).
 FT METAL 16 16 IRON (HEME AXIAL LIGAND).
 FT METAL 66 66 IRON (HEME AXIAL LIGAND).
 FT SEQUENCE 92 AA; 9781 MW; 573C3CF0 CRC32;
 CY2_RHOTE Length: 92 February 14, 2000 08:02 Type: P Check: 2620 ..
 1 ADESLAQTK GCLACHNPEK KVVGPAYGVY AKRYAGAGA EAKLVAKVNA
 51 GGQGVMAKOL GAEIPYANN VTKREATLV KWTLSKQID YK
 11AA SEQUENCE 1.0 STANDARD; PRT; 299 AA.
 ID CYNR.ECOLI
 AC P27111.
 DT 01-FEB-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE CYN OPERON TRANSCRIPTIONAL ACTIVATOR.
 GN CYN.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12.
 RX MEDLINE: 92276346.
 RA SUNG Y.-C., FUCHS J.A.;
 RT "The Escherichia coli K-12 cyn operon is positively regulated by a member of the *lyrA* family";
 RT J. Bacteriol. 174:3645-3650(1992).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE: 97426617.
 RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNAN N.T., BURLAND V., RILEY M., COLLADO-VIDES J., GLASNER F.D., ROBE C.K., MAYHEM G.F., GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J., MAU B., SHAO Y.;
 RT "The complete genome sequence of Escherichia coli K-12";
 RT Science 277:1453-1474(1997).
 RL [3]
 RP SEQUENCE FROM N.A.
 RC DUNCAN M., ALLEN E., ARAUJO R., APARICIO A.M., CHUNG E., DAVIS K., FERREPEITL N., HYMAN R., KALMAN S., KOMP C., KURI O., LEW H., LIN D., NAMATH A., OEFNER P., ROBERTS D., SCHRAM S., DAVIS R.W.;
 RA Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
 RL FUNCTION: POSITIVE REGULATORY PROTEIN FOR THE CYN OPERON.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: BELONGS TO THE *LYR* FAMILY OF TRANSCRIPTIONAL REGULATORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M93053; AA23628.1; ..
 DR EMBL: AE000141; AAC73441.1; ..
 DR EMBL: U73857; AAB18062.1; ALT_INIT.
 DR PIR: A41900; A41900.

ECOGENE: EG11421; CYNR.
 DR PROSITE: P500044; HTL_LYR_FAMILY; 1.
 DR PRAM: PF00126; HTL_1; 1.
 KW Transcription regulation; DNA-binding; Activator.
 FT DNABIND 18 37 H-T-H motif (BY SIMILARITY).
 FT SEQUENCE 299 AA; 32961 MW; 0A8298D6 CRC32;
 CYNR.ECOLI Length: 299 February 14, 2000 08:02 Type: P Check: 4304 ..
 1 MSTRINFL AVAHSSEFR AASALHSQP ALSQIQROLE ESLGVPLEDR
 51 SGRTRLIDA GEWRQYASR ALQELGAKR AIHDVADLR GSLRIAVTPT
 101 FTSYFGLPM ADFARYPSI TLQLOMSQE KIDMLCRDE LDVGIAFAPV
 151 HSPLEALPL LTESLAHYA QHHPPLAVHQ VALSRLHDEK IYVLSAEFAT
 201 REQIDHYCEK AGLHPQVTE ANSISVLEL IRRSTSLTL PAIAFQHDG
 251 LKAISLAPPL LERTAVLIR KNSWQTAAK AFLHMAIDKC AVAGNESR
 11AA SEQUENCE 1.0 STANDARD; PRT; 291 AA.
 ID DAXX_STRSK
 AC P39042;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE D-ALANYL-D-ALANINE CARBOXYPEPTIDASE PRECURSOR (EC 3.4.16.4) (DD-PEPTIDASE) (DD-CARBOXYPEPTIDASE).
 GN Streptomyces sp. (strain K15).
 OS Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Streptomyces.
 OC Actinomycetales; Streptomycetales; Streptomycetaceae; Streptomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE: 92028812.
 RA PALOMER-DURSTCHE M., DIEZ C., HOUBA S., DIBBERG O., VAN BREDEN J., GHUYSEN J.-M.;
 RT "Amino acid sequence of the penicillin-binding protein/penicillin-binding proteins of class A";
 RT Biochem. J. 279:223-230(1991).
 RL [2]
 RP SEQUENCE OF 30-85, AND ACTIVE SITE.
 RC MEDLINE: 89350853.
 RA LEYH-BOUILLE M., VAN BREDEN J., RENIER-PRIOT S., JORIS B., NGUYEN-DISTECHE M., GHUYSEN J.-M.;
 RT "The streptomyces K15 DD-peptidase/penicillin-binding protein. Active site and sequence of the N-terminal region";
 RT Biochem. J. 260:601-604(1989).
 RL [3]
 RP FUNCTION: REMOVES C-TERMINAL D-ALANYL RESIDUES FROM SUGAR-PEPTIDE
 CC -1- CELL WALL PRECURSORS.
 CC -1- CATALYTIC ACTIVITY: D-ALANYL-D-ALANINE + H₂O = 2 D-ALANINE.
 CC -1- PATHWAY: FINAL STAGES IN PEPTIDOGLYCAN SYNTHESIS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S11. ALSO KNOWN AS THE D-ALANYL-D-ALANINE CARBOXYPEPTIDASE 1 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X59965; CAA42591.1; ..
 DR PIR: S04638; S04638.
 DR PIR: S17674; S17674.
 DR PRAM: PF00768; Peptidase_S11; 1.
 KW Hydrolyase; Carboxypeptidase; Peptidoglycan synthesis; Cell wall; Signal.
 FT SIGNAL 1 29
 FT CHAIN 30 291 D-ALANYL-D-ALANINE CARBOXYPEPTIDASE.

ACT SITE 64 64 ACYLATED BY PENICILLIN.
 FT 70 T -> Q (IN REF. 2).
 FT 83 D -> S (IN REF. 2).
 SO SEQUENCE 291 AA; 30257 MW; E2D9DFE9 CRC32;

DACC_STRSK Length: 291 February 14, 2000 08:02 Type: P Check: 4817

1 MRLRAATV ITTGALLAG TLGATPATAV TKPTIAVGC YAMNNGTGT
 51 LYTAADTRR STGSTRTMT AKVLAOSNL NLDKVTICR AYSDYVANK
 101 PSOHLIVGD KTVROLLYG LMLPSGCDAA YALADKXSG SOAARVASF
 151 IGMNTAATN LGLNTHFDS FDGIGNGANY STPRHLTKIA SSAMKSTFR
 201 TVVTKAYTA KTVTKGSIR TMDTKWNTNG LLSYSGLAIG VKTSGPEAK
 251 YCLVPAATRG KTVIGTVLA STSIPARESD ATKIMYGA L

11AA_SEQUENCE 1.0
 ID DBPS_YEAST STANDARD; PRT: 482 AA.
 AC P20449;

DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DE 15-JUL-1998 (Rel. 36, Last annotation update)
 GN PROBABLE ATP-DEPENDENT RNA HELICASE DBPS (HELICASE CA5/6).
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA CHANG T.-H.;
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RA LANDT O., HIESER R., UNSEID M.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-247 FROM N.A.
 RA BOHN C., BOLOTIN-FUKUHARA M., DAIGMAN-FORMIER B., DANG D.V.,
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 237-425 FROM N.A.
 RA CHANG T.-H., ARENAS J., ABELSON J.;
 RL Identical to five putative yeast RNA helicase genes.;

CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
 CC -1- SIMILARITY: TO OTHER "DEAD" BOX FAMILY HELICASES
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

EMBL; U28135; AAB01679.1;
 EMBL; 274954; CAAG9237.1;
 EMBL; 274955; CAAG9239.1;
 PIR; D34848; D34848;
 SGD; L0003292; DBPS.
 DR PROSITE: PS00039; DEAD_ATP_HELICASE; 1.
 DR PFAM: PF00270; DEAD.1.
 DR PFAM: PF00271; helicase.C; 1.
 NP_BIND; 138 145 ATP (POTENTIAL).
 FT 239 242 DEAD BOX.
 FT 482 AA; 53874 MW; C70FB665 CRC32;

DBPS_YEAST Length: 482 February 14, 2000 08:02 Type: P Check: 2958

1 MSDTRKPAD LKSLKIDNE KEDTSEVSTR ETVKSQPEKT ADSTKPAEKL
 51 VPKVEERKTR QEDSNLISE YEVKVLADI QADPNSEPLS AKSDDELGLA
 101 PELKGIYAM KFOKPSKIOE RALPLILHP PNMIAOSQS GTGTAEPFL
 151 TMLTRVPED ASPOAICLAP SRELARQTL VQEGKFLK ITSQLIPDS
 201 FEKNKOINQ VIVGPGTVL DLMRKLMQ OKIKFVLDE ADNMDOOGL
 251 GDOCIKRFK LPRDQVLF SATFADAVNG YAKKIYVPAK TLELQNEVN
 301 VDAIKOLYMD CKNEADKFDV LTELXGLMT GSSITFVAK KIANVLYGKL
 351 KSEGHEVSL HGDQTOERD RLIDDPREG SKVLITNVL ARGIDIPVVS
 401 MYVNDLPTL ANQADAPATY IHRIGTRGF GRKVALSEV HDKNSENILS
 451 AIOKFGDIE MTRVPTDWD EYEKIVKVL KD

11AA_SEQUENCE 1.0
 ID DBP_STRMU STANDARD; PRT: 145 AA.
 AC Q54433;

DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE DNA/PANTOTHENATE METABOLISM FLAVOPROTEIN HOMOLOG (FRAGMENT).
 GN DBP
 OS Streptococcus mutans.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA GUTIERREZ J.A., CROWLEY P.J., BROWN D.P., HILLMAN J.D.,
 RX MEDLINE; 96312359.
 RA YOUNGMAN P., BLEIWEIS A.S.;
 RT "Insertional mutagenesis and recovery of interrupted genes of
 RT Streptococcus mutans by using transposon Tn917: preliminary
 RT characterization of mutants displaying acid sensitivity and
 RT nutritional requirements.";
 RL J. Bacteriol. 178:4166-4175(1996).
 CC -1- FUNCTION: FLAVOPROTEIN AFFECTING SYNTHESIS OF DNA AND PANTOTHENATE
 CC METABOLISM (BY SIMILARITY).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

EMBL; U48885; AAC44502.1;
 FT NON_TER 1
 FT NON_TER 145
 SO SEQUENCE 145 AA; 15819 MW; 24EB4944 CRC32;

DPP_STRMU Length: 145 February 14, 2000 08:02 Type: P Check: 9135

1 MTKKILLAVS GSIAAYKAD ISHQITKIGY HNVNFMNAA KQFIPPLTQ
 51 VLSKNPVYSN VKMEDDPOVI NHIALAKOAD LFLPPASAN TLALAHGHA
 101 DNIVTSVALA LPLEVKEFA PAMNITKVEN PIQNSNTLL KKFYG

11AA_SEQUENCE 1.0
 ID DBAT_CITER STANDARD; PRT: 387 AA.
 AC P45513;
 DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, last sequence update)
 DT 01-NOV-1997 (Rel. 35, last annotation update)
 DE 1,3-PROPANEDIOL DEHYDROGENASE (EC 1.1.1.202) (3-HYDROXYPROPYIONALDEHYDE
 DE REDUCTASE) (1,3-PROPANEDIOL OXIDOREDUCTASE).
 GN DHAT.
 OS Citrobacter freundii.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Citrobacter.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DSM 30040;
 RX MEDLINE; 95238288.
 RA DANIEL R., BOENIGK R., GOTTSCHEK G.;
 RT Purification of 1,3-propanediol dehydrogenase from Citrobacter
 RT freundii and cloning, sequencing, and overexpression of the
 RT corresponding gene in Escherichia coli.";
 RL J. Bacteriol. 177:2151-2156(1995).
 CC -1- FUNCTION: MOST ACTIVE WITH SUBSTRATES CONTAINING TWO PRIMARY
 CC ALCOHOL GROUPS SEPARATED BY ONE OR TWO CARBON ATOMS. IN THE
 CC PHYSIOLOGICAL DIRECTION, 3-HYDROXYPROPYIONALDEHYDE IS THE PREFERRED
 CC SUBSTRATE.
 CC -1- CATALYTIC ACTIVITY: PROPANE-1,3-DIOL + NAD(+) = 3-HYDROXYPROPANAL
 CC + NADH.
 CC -1- COFACTOR: IRON.
 CC -1- SUBUNIT: HOMOCYMER.
 CC -1- SIMILARITY: BELONGS TO THE IRON-CONTAINING ALCOHOL DEHYDROGENASE
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; 009771; AAB48848.1;
 CC PROSITE; P500913; ADH_IRON_1; 1.
 CC DR PROSITE; P500913; ADH_IRON_2; 1.
 CC DR PFAM; PF00465; Fe-ADH; 1.
 CC KX Oxidoreductase; NAD; Iron.
 CC KM Oxidoreductase; NAD; Iron.
 CC SQ SEQUENCE 387 AA; 41481 MW; BAC88846 CRC32;
 DHAT_CITFR Length: 387 February 14, 2000 08:02 Type: P Check: 101 ..
 ID MSYRMEDYLV PNVNEFGPNA ISVYGERCKL LGKKKALLVY DKGRLAKDQ
 1
 51 AVDKTLHLR EAGIDVVFED GVEPNKDTN VRDGLVFRK EECDDIVTVG
 101 GGSFHDGCKG IGIAATHEGD LYSVAGIELT TNPPLPIYAV NTAAGTASEV
 151 TRHCVLNTR TKRKFVIVSW RNLPSVIND PLMLGKPA LRAATGMAL
 201 THAVEATISK DANPYTDAAA IQAIRLIARN LROAVALGNS LQARENMAVA
 251 SLGAGAFNN ANLGYVHAMA HOLGLYDMP HGVANAVILP HVARYNLIAN
 301 PEKADIAEL MGENITGLST MDAELAIAH IARLSADIGI POHLRLDGVK
 351 EADFPYAEW ALKDGNAFSN PRKNGEKEIA EIFROAF
 IDA SEQUENCE 1.0 STANDARD: PRT; 387 AA.
 ID DHAT_KLEPN
 AC 059477;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, last sequence update)
 DT 01-NOV-1997 (Rel. 35, last annotation update)
 DE 1,3-PROPANEDIOL DEHYDROGENASE (EC 1.1.1.202) (3-HYDROXYPROPYIONALDEHYDE
 DE REDUCTASE) (1,3-PROPANEDIOL OXIDOREDUCTASE).
 GN DHAT.
 OS Klebsiella pneumoniae.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Klebsiella.

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 25955;
 RA SKRAI F.A., WILLARD B.L., CAMERON D.C.;
 RL Submitted (NOV-1995) to the EMBL/Genbank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: PROPANE-1,3-DIOL + NAD(+) = 3-HYDROXYPROPANAL
 CC + NADH.
 CC -1- COFACTOR: IRON (BY SIMILARITY).
 CC -1- SUBUNIT: HOMOCYMER (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE IRON-CONTAINING ALCOHOL DEHYDROGENASE
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; 030903; AAA74260.1;
 CC PROSITE; P500913; ADH_IRON_1; 1.
 CC DR PROSITE; P500913; ADH_IRON_2; 1.
 CC DR PFAM; PF00465; Fe-ADH; 1.
 CC KM Oxidoreductase; NAD; Iron.
 CC SQ SEQUENCE 387 AA; 41465 MW; 04B25CC1 CRC32;
 DHAT_KLEPN Length: 387 February 14, 2000 08:02 Type: P Check: 846 ..
 ID MSYRMEDYLV PNVNEFGPNA ISVYGERCKL LGKKKALLVY DKGRLAKDQ
 1
 51 AVDKTLHLR EAGIDVVFED GVEPNKDTN VRDGLVFRK EECDDIVTVG
 101 GGSFHDGCKG IGIAATHEGD LYSVAGIELT TNPPLPIYAV NTAAGTASEV
 151 TRHCVLNTR TKRKFVIVSW RNLPSVIND PLMLGKPA LRAATGMAL
 201 THAVEATISK DANPYTDAAA IQAIRLIARN LROAVALGNS LQARENMAVA
 251 SLGAGAFNN ANLGYVHAMA HOLGLYDMP HGVANAVILP HVARYNLIAN
 301 PEKADIAEL MGENITGLST LQAEKAIAH IRLSMIDIGI POHLRLDGVK
 351 EADFPYAEW ALKDGNAFSN PRKNGEKEIA AIFROAF
 IDA SEQUENCE 1.0 STANDARD: PRT; 310 AA.
 ID DHAT_HUMAN
 AC P37058;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-NOV-1995 (Rel. 32, last sequence update)
 DT 15-DEC-1998 (Rel. 37, last annotation update)
 DE ESTRADIOL 17-BETA-DEHYDROGENASE 3 (EC 1.1.1.62) (17-BETA-HSD 3)
 DE (TESTICULAR 17-BETA-HYDROXYSTEROID DEHYDROGENASE).
 GN HSD17B3 OR EDH1/B3.
 OS Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 CC [1]
 CC SEQUENCE FROM N.A., AND VARIANTS MPH Q-80; V-203; K-232 AND V-235.
 CC TISSUE-TESTIS;
 RC MEDLINE; 94355972.
 RA GEISLER W.M., DAVIS D.L., WU L., BRADSHAW K.D., PAEL S.,
 RA MENDONCA B.B., ELLISTON K.O., WILSON J.D., RUSSELL D.W.,
 RA ANDERSSON S.;
 RT "Male pseudohermaphroditism caused by mutations of testicular 17
 RT beta-hydroxysteroid dehydrogenase 3.";
 RL Nat. Genet. 7:34-39(1994).
 CC [2]
 CC VARIANT MPH TRP-80.
 CC MEDLINE; 98429412.
 RA BILBAO J.R., LORIDAN L., AUDI L., GONZALO E., CASTANO L.;
 RT "A novel missense (R80W) mutation in 17-beta-hydroxysteroid
 RT dehydrogenase type 3 gene associated with male

RT Pseudohermaphroditism.
 RT Eut. J. Endocrinol. 139:330-333(1998).
 CC -1- FUNCTION: FAVORS THE REDUCTION OF ANDROSTENEDIONE TO TESTOSTERONE.
 CC -1- USES NADPH WHILE THE TWO OTHERS EDH17B ENZYMES USE NADH.
 CC -1- CATALYTIC ACTIVITY: ESTRADIOL-17-BETA + NADP(+) = ESTRONE + NADPH.
 CC -1- PATHWAY: LAST ENZYME OF THE TESTOSTERONE BIOSYNTHESIS.
 CC -1- DISEASE: DEFECTS IN EDH17B GIVE RISE TO MALE PSEUDHERMAPHRODISM (MPH). THESE 46 XY INDIVIDUALS HAVE UNMIBIGIOUS FEMALE EXTERNAL GENITALIA AT BIRTH, BUT FAIL TO MENSTRUATE AT THE TIME OF EXPECTED PUBERTY AND INSTEAD VIRILISE AS EVIDENCED BY GROWTH OF THE PHALLUS. BREAST DEVELOPMENT MAY OR MAY NOT TAKE PLACE.
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES FAMILY (SDR). 17-BETA-HSD 3 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U05659; AAC50066.1;
 DR MIM: 264300;
 DR PROSITE: PS00061; ADH_SHORT: 1.
 DR PFM: PFM0106; adh_short: 1.
 KW Steroid biosynthesis; Oxidoreductase; NADP; Multigene family;
 KW Pseudohermaphroditism; Disease mutation.
 FT NP_BIND 48 NADP (BY SIMILARITY).
 FT ACT_SITE 198 198 R -> Q (IN MPH; GAZA).
 FT VARIANT 80 80 R -> Q (IN MPH; GAZA).
 FT VARIANT 80 80 R -> W (IN MPH).
 FT VARIANT 203 203 A -> V (IN MPH).
 FT VARIANT 232 232 S -> L (IN MPH).
 FT VARIANT 235 235 M -> V (IN MPH).
 FT SEQUENCE 310 AA; 34515 MW; 652F1024 CRC32;
 DHB3_HUMAN Length: 310 February 14, 2000 08:02 Type: P Check: 3434 ..
 1 MGDVLEQFEFI LTGLVCLAC LACVRFSRC VLLNWKVLP KSLRLRSGGV
 51 AVITGAGDGI GKAYSFELAK RGLNVVLISR TLEKLEAIAT EIERITGRSV
 101 KIIOADFTKD DIVEHIREKL AGLEIGILVN NVGMLPNLTP SHLNPDEI
 151 OSLHCNITS VKMTOLILK HMESSROKGLI LNISSGIALF PMLYIMSYSA
 201 SKAFVCASFQ ALOEYKAKE VLIQVLTPEYA VSTAMTKYLN TNYITTADE
 251 FVKESLINVT IGGETGCCIA HEILAGFSL IPANAFYSGA FQRLILTHYV
 301 AYLRNLTQVR
 11AA_SEQUENCE 1.0
 ID DHB3_MOUSE STANDARD; PRT; 305 AA.
 AC P70365;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE (TESTICULAR 17-BETA-DEHYDROGENASE 3 (EC 1.1.1.62) (17-BETA-HSD 3)
 DE (TESTICULAR 17-BETA-HIDROXISTEROID DEHYDROGENASE).
 OS Mus musculus (Mouse)
 GN HSD17B3 OR EDH17B3.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.

RC TISSUE-TESTIS;
 RA MEDLINE; 97325842.
 RA SHA J.A., DUDLEY K., RAJAPAKSHA W.R.A., O'SHAUGHNESSY P.J.;
 RT "Sequence of mouse 17beta-hydroxysteroid dehydrogenase type 3 cDNA
 RT and tissue distribution of the type 1 and type 3 isoform mRNAs".
 RT J. Steroid Biochem. Mol. Biol. 60:19-24(1997).
 CC -1- FUNCTION: FAVORS THE REDUCTION OF ANDROSTENEDIONE TO TESTOSTERONE.
 CC -1- USES NADPH WHILE THE TWO OTHERS EDH17B ENZYMES USE NADH (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ESTRADIOL-17-BETA + NADP(+) = ESTRONE + NADPH.
 CC -1- PATHWAY: LAST ENZYME OF THE TESTOSTERONE BIOSYNTHESIS.
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES FAMILY (SDR). 17-BETA-HSD 3 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U66827; AAB06793.1;
 DR MGD: MG1:107177; HSD17B3.
 DR PROSITE: PS00061; ADH_SHORT: 1.
 DR PFM: PFM0106; adh_short: 1.
 KW Steroid biosynthesis; Oxidoreductase; NADP; Multigene family.
 FT NP_BIND 44 73 NADP (BY SIMILARITY).
 FT ACT_SITE 194 194 BY SIMILARITY.
 FT SEQUENCE 305 AA; 34271 MW; 10CE2855 CRC32;
 DHB3_MOUSE Length: 305 February 14, 2000 08:02 Type: P Check: 9317 ..
 1 MERFLIAGL FVGLVCLVKC MRESQLFLR FCKALPSSFL RSMGQNAVIT
 51 GAGDIGRNY SFELARGLN VYLISRTLEK LQIAELIER TIGSGYKIYQ
 101 ADFTREDIYD HIKHEGLEGL NGLVNVNNGM LPSTFPBSHL SSGESQNI
 151 HONITSYVKM TQVLKHMES RRGGLINIS SGALRBPWL YSLYSASNAF
 201 VTFESKALSV EYRDKGIIID VLPYISITP MTKYLNKKMT KTADEFVKS
 251 LKYVIGAES CGCLAHETIA IILNRPISRI FYSTAGRELT LTRYSDYLKR
 301 NISNR
 11AA_SEQUENCE 1.0
 ID DHB3_RAT STANDARD; PRT; 306 AA.
 AC O54939;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE (TESTICULAR 17-BETA-DEHYDROGENASE 3 (EC 1.1.1.62) (17-BETA-HSD 3)
 DE (TESTICULAR 17-BETA-HIDROXISTEROID DEHYDROGENASE).
 OS Rattus norvegicus (Rat)
 GN HSD17B3 OR EDH17B3.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA KHANUM A., TSAI-MORRIS C.-H., DUFAU M.L.;
 RT Submitted (Nov-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: FAVORS THE REDUCTION OF ANDROSTENEDIONE TO TESTOSTERONE.
 CC -1- USES NADPH WHILE THE TWO OTHERS EDH17B ENZYMES USE NADH (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ESTRADIOL-17-BETA + NADP(+) = ESTRONE + NADPH.
 CC -1- PATHWAY: LAST ENZYME OF THE TESTOSTERONE BIOSYNTHESIS.
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES FAMILY (SDR). 17-BETA-HSD 3 SUBFAMILY.
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: AF035156; AB99739.1; FALSE_NEG.
 PROSITE: PS00061; ADH_SHORT; 1.
 PFAM: PF00106; adh_short; 1.
 Steroid biosynthesis; Oxidoreductase; NADP; Multigene family.
 NP_BIND 44 194 BY SIMILARITY.
 ACT_SITE 194 194 BY SIMILARITY.
 SEQUENCE 306 AA: 34223 MW: 34223 MW: BFBBCB99 CRC32:
 DHX3_RAT Length: 306 February 14, 2000 08:02 Type: P Check: 3366

1 MEQFLSLVGL LVCLVCLVKC VFRSRYFLFS FCKALPGSFL RSMGQNAVIT
 51 GAGDGLKAY SPELARHGLN VLIISRLK LQVISEIER TGSRRKVVQ
 101 ADFTREDITD HIEBQKGL EGVLVNNGM IPNLPSHFL STGESQSVI
 151 HCNITSVYKM TOLVKNHES RRGGLINIS SGVGRPMPL YSLYSKRAF
 201 VCFESKALNV EYRDKGIITQ VLPYVSVP MTKYINTSRV TKTADEFVKE
 251 SKTYVTIGAE TCGCLAHETL AILNLPSR IYSSSTORF LNKOFSDYK
 301 SNISNR

11AA_SEQUENCE 1.0 STANDARD: PRT: 312 AA.
 ID DHX3_MOUSE
 AC 070503;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE PUTATIVE STEROID DEHYDROGENASE KIK-1 (EC 1.1.1.-).
 GN KIK1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPRAIN-BALB/C; TISSUE-LIVER;
 RA GAMBORTO A., PAGLIANO O., ROBBINS P., DELBO A.;
 RL Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.
 CC -1 SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 CC -1 FAMILY (SDR). 17-BETA-HSD 3 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL: AF064635; AAC1685.1;
 PROSITE: PS00061; ADH_SHORT; 1.
 PFAM: PF00106; adh_short; 1.
 Steroid biosynthesis; Oxidoreductase; NADP; Multigene family.
 NP_BIND 50 202 BY SIMILARITY.
 ACT_SITE 202 202 BY SIMILARITY.
 SEQUENCE 312 AA: 34742 MW: 12162554 CRC32:
 DHX3_MOUSE Length: 312 February 14, 2000 08:02 Type: P Check: 4788

1 MECAPPAAGF LYWGASTIA YLALRSYSL FRAVWCYGV NEALVGPRLG
 51 EMAYVTGTD GIKAYAEEL AKGAKIVLI SRQDKLNOV SNRIKEKRV

101 ERTIADVDS LDDIYDKRT GISGLIEGLV YNNGMSYEV PEYFLPIPLD
 151 DNTIKLINI NVLSVCKVTR LVLPWVRS KVIINISSA SGMPLVPLLT
 201 IYATKAFVD FFSQCLHEEY KSKGIFVOS MPLVATKLA KIQPPLDKP
 251 SATFEKSAI KTVGLQTRTT GYVISHLMS INSIMPRMW FKINIGFESK
 301 LRRNRLKRR KN

11AA_SEQUENCE 1.0 STANDARD: PRT: 312 AA.
 ID DHX3_HUMAN
 AC 075828;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE PUTATIVE STEROID DEHYDROGENASE SPW2 (EC 1.1.1.-).
 GN ANAS platyrhynchos (Domestic duck).
 OS Anas platyrhynchos (Domestic duck).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves; Eukaryota; Metazoa; Chordata; Anatidae; Anas.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-FOREBRAIN;
 RA KIMURA N., KUROSAWA N., KONDO K., TSUKADA Y.;
 RL Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.
 CC -1 TISSUE SPECIFICITY: BRAIN.
 CC -1 SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 CC -1 FAMILY (SDR). 17-BETA-HSD 3 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL: AB009304; BAA23765.1;
 PROSITE: PS00061; ADH_SHORT; 1.
 PFAM: PF00106; adh_short; 1.
 Steroid biosynthesis; Oxidoreductase; NADP; Multigene family.
 NP_BIND 48 201 BY SIMILARITY.
 ACT_SITE 201 201 BY SIMILARITY.
 SEQUENCE 312 AA: 33896 MW: 3DBH3A6D CRC32:
 DHX3_ANAPL Length: 312 February 14, 2000 08:02 Type: P Check: 9157

1 MPRAGILMW VGALGALYAA VRGALGILGA LRVWGIGAGR AALGPGLGN
 51 AVYGTATGCI GKAYAKELAK RGMVVALISR SKKLDVYAG EITBQYVER
 101 KYIVADFGER EDIYDLRAG LEGLEIGLV NNVGISYSYP EYFIDVDDL
 151 KTIQKININ IMVCKMTRV VLPGLMERSK GVLINISSA GMYPTPLILT
 201 YSASKAFVDY FSRDLHAEY SKGIIVQSVM PYYVATKMSK ISKPSDKPT
 251 PETYVAIGA TVGLQSTING CLPHAFMGV FSILPTIYV NLMTKTKOI
 301 RARFLKRRK EK

11AA_SEQUENCE 1.0 STANDARD: PRT: 276 AA.
 ID DHX3_HUMAN
 AC 075828;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 39, Last annotation update)
 DE CARBONYL REDUCTASE [NADPH] 3 (EC 1.1.1.184) (NADPH-DEPENDENT CARBONYL
 DE REDUCTASE 3).
 GN Homo sapiens (Human).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 98414514.
 RA WATANABE K., SUGAMARA C., ONO A., FUKUZUMI Y., ITAKURA S.,
 RA YAMAZAKI M., TASHIRO H., OSEOGAWA K., SOEDA E., NOMURA T.;
 RT "Mapping of a novel human carbonyl reductase, CBR3, and ribosomal
 RT pseudogenes to human chromosome 21q22.2.";
 RT Genomics 52:95-100(1998).
 CC -1- CATALYTIC ACTIVITY: R-CHOH-R' + NADP(+) = R-CO-R' + NADPH.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 CC FAMILY (SDR).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AB004854; BAA33500.1; -;
 DR EMBL: AB003151; BAA34207.1; -;
 DR MIM: 603608; -;
 DR PROSITE: PS00061; ADH_SHORT; 1.
 DR PFM: PFO0106; adh_short; 1.
 KW Oxidoreductase; NADP.
 FT INIT_MET 0 0 BY SIMILARITY.
 FT NE_BIND 9 33 NADP (BY SIMILARITY).
 FT ACT_SITE 193 193 BY SIMILARITY.
 SQ SEQUENCE 276 AA; 30719 MW; AETD716B CRC32;
 DHCA_HUMAN Length: 276 February 14, 2000 08:02 Type: P Check: 391 ..
 1 SSGSRVALVT GANRGIGLAI ARELCROFSG DVLTRADVA RGQAAYOQLQ
 51 AEGLSPRFHQ LDIDDLQISIR ALRDLRKEY GGLNLYVNNA AVAKSDPPM
 101 PFDIKAMTL KTNFATRNK CNELIPIMK HGRVYNISL QCLRAFENCS
 151 EDQERFHE TLTEGLDVL MKRVEDTKN EVHERGWN SPYSGKLG
 201 TVLSRIARR LDKRRADRI LVNACCGPV KIDMGKXSI RTVBEGMETP
 251 VYIALLEPDA TEPQGLYHD KYVQNM
 1AA_SEQUENCE 1.0 STANDARD; PRT; 276 AA.
 ID DHCA_HUMAN
 AC P16152;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE CARBOXYL REDUCTASE [NADPH] 1 (EC 1.1.1.184) (NADPH-DEPENDENT CARBOXYL
 DE REDUCTASE 1) (PROSTAGLANDIN-E2 9-REDUCTASE) (EC 1.1.1.189)
 DE (PROSTAGLANDIN 9-KETOREDUCTASE) (15-HYDROXYPROSTAGLANDIN DEHYDROGENASE
 DE (NADP+)) (EC 1.1.1.197).
 GN CBR1 OR CBR OR CRN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE-PLACENTA;
 RX MEDLINE: 89034082.
 RA WERNITZ B., BOHREN K.M., HEINEMANN G., VON WARTBURG J.-P.,
 RA GABBAY K.H.;
 RT "Human carbonyl reductase. Nucleotide sequence analysis of a cDNA and
 RT amino acid sequence of the encoded protein.";
 RL J. Biol. Chem. 263:16185-16188(1988).
 RN [2]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE-BREAST;
 RX MEDLINE: 90212644.

RA FORREST G.L., AKMAN S., KRUTZIK S., PAXTON R.J., SPARKES R.S.,
 RA DOROSHOW J., FELSTED R.L., MOHANDAS T., BACHUR N.R.;
 RT "Induction of a human carbonyl reductase gene located on chromosome
 RT 21.";
 RL Biochim. Biophys. Acta 1048:149-155(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 92017676.
 RA FORREST G.L., AKMAN S., DOROSHOW J., RIVERA H., KAPLAN W.D.;
 RT "Genomic sequence and expression of a cloned human carbonyl reductase
 RT gene with daunorubicin reductase activity.";
 RL Mol. Pharmacol. 40:502-507(1991).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 98414514.
 RA WATANABE K., SUGAMARA C., ONO A., FUKUZUMI Y., ITAKURA S.,
 RA YAMAZAKI M., TASHIRO H., OSEOGAWA K., SOEDA E., NOMURA T.;
 RT "Mapping of a novel human carbonyl reductase, CBR3, and ribosomal
 RT pseudogenes to human chromosome 21q22.2.";
 RL Genomics 52:95-100(1998).
 RN [5]
 RP PARTIAL SEQUENCE, AND POST-TRANSLATIONAL MODIFICATIONS.
 RX MEDLINE: 9313816.
 RA KROOK M., GHOSH D., STROEMBERG R., CARLOUSET M., JOERNVALL H.;
 RT "Carboxylethyllysine in a protein: native carbonyl reductase/NADP(+) -
 RT dependent prostaglandin dehydrogenase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:502-506(1993).
 CC -1- FUNCTION: CATALYZE THE REDUCTION OF A WIDE VARIETY OF CARBOXYL
 CC COMPOUNDS INCLUDING THE ANTITUMOR ANTHRACYCLINE ANTIBIOTICS.
 CC CAN CONVERT PROSTAGLANDIN E2 TO PROSTAGLANDIN F2-ALPHA.
 CC -1- CATALYTIC ACTIVITY: R-CHOH-R' + NADP(+) = R-CO-R' + NADPH.
 CC -1- CATALYTIC ACTIVITY: (5Z,13E)-(15S)-9,11,15-TRIHYDROXYPROSTA-5,13-
 CC DIENATE + NADP(+) = (5Z,13E)-(15S)-11,15-DIHYDROXY-9-OXOPROSTA-
 CC 5,13-DIENATE + NADPH.
 CC -1- CATALYTIC ACTIVITY: (5Z,13E)-(15S)-11-ALPHA,15-DIHYDROXY-9-
 CC OXOPROST-13-ENATE + NADP(+) = (5Z,13E)-11-ALPHA-HYDROXY-9,15-
 CC DIOXOPROST-13-ENATE + NADPH.
 CC -1- SUBUNIT: MONOMER.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 CC FAMILY (SDR).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: J04056; AAA52070.1; -;
 DR EMBL: M62420; AAA17881.1; -;
 DR EMBL: AB003151; BAA33498.1; -;
 DR PIR: A31912; RDHUBC.
 DR PIR: S09013; S09013.
 DR HSSP: P14061; 1PDM.
 DR MIM: 114830; -;
 DR PROSITE: PS00061; ADH_SHORT; 1.
 DR PFM: PFO0106; adh_short; 1.
 KW Oxidoreductase; NADP; Acetylation.
 FT INIT_MET 0 0
 FT MOD_RES 1 1 ACETYLATION.
 FT NE_BIND 238 238 N6-(1-CARBOXYETHYL).
 FT ACT_SITE 9 33 NADP (BY SIMILARITY).
 FT ACT_SITE 193 193 BY SIMILARITY.
 SQ SEQUENCE 276 AA; 30244 MW; 470A0F07 CRC32;
 DHCA_HUMAN Length: 276 February 14, 2000 08:02 Type: P Check: 3202 ..
 1 SSGIRVALVT GGNKGIGLAI VRDLCRLFSG DVLTRADVT RGQAAYOQLQ
 51 AEGLSPRFHQ LDIDDLQISIR ALRDLRKEY GGLNLYVNNA GIAKRVADPT


```

101 PHIOAEVTM KTFEGTRDV CTELLPLIK QGRVNVSSSI MSVALKSCS
151 PELQKFRSE TITEELVGL MKNFVEDTK GVHKEGSPS SAICVTIKIGV
201 TVLSRIHARK LSEQRKGDKI LNVACCPGV RTDNAGPRAT KSPBEGAETP
251 VYIALPPDA EGPHQGFVSD KRYEQM

11AA_SEQUENCE 1.0
ID DHCA_MOUSE STANDARD: PRT: 276 AA.
AC P47727;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 15-FEB-1996 (Rel. 36, Last annotation update)
DE CARBONYL REDUCTASE [NADPH] (EC 1.1.1.184) (NADPH-DEPENDENT CARBONYL
REDUCTASE).
GN CBR OR CBR1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN (1)
RP SEQUENCE FROM N.A.
RA MEDLINE: 9629674.
RA WEI J., DLOUTH S.R., HARA A., GHETTI B., HODES M.E.;
RT Cloning a cDNA for carbonyl reductase (CBR) from mouse cerebellum:
RT murine genes that express cbr map to chromosomes 16 and 11."
RL Genomics 34:147-148(1996).
CC -1- CATALYTIC ACTIVITY: R-CHOH-R' + NADPH(+) -> R-CO-R' + NADPH.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
FAMILY (SDR).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U31666; AAB19006.1; -
DR DGD; MG: 88284; CBR.
DR PROSITE: P800061; ADH_SHORT; 1.
DR PFAM: PF00106; adh_short; 1.
KW Oxidoreductase; NADP; Acetylation.
FT INIT_MET 0 BY SIMILARITY.
FT MOD_RES 1 ACETYLATION (BY SIMILARITY).
FT NP_BIND 9 NADP (BY SIMILARITY).
FT ACT_SITE 193 BY SIMILARITY.
FT SEQUENCE 276 AA; 30597 MW; 09438E74 CRC32;

DHCA_MOUSE Length: 276 February 14, 2000 08:02 Type: P Check: 204

```

```

DE CARBONYL REDUCTASE [NADPH] (EC 1.1.1.184) (NADPH-DEPENDENT CARBONYL
REDUCTASE).
GN CBR.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN (1)
RP SEQUENCE FROM N.A.
RA MEDLINE: 95197021.
RA GONZALES B., SAPIA A., RIVERA H., KAPLAN W.D., YAM B., FORREST G.L.;
RT Cloning and expression of the cDNA encoding rabbit liver carbonyl
RT reductase."
RL Gene 154:297-298(1995).
CC -1- FUNCTION: CATALYZES THE REDUCTION OF A WIDE VARIETY OF CARBONYL
CC COMPOUNDS INCLUDING THE ANTITUMOR ANTIRACICLINE ANTIBIOTICS.
CC -1- CATALYTIC ACTIVITY: R-CHOH-R' + NADPH(+) -> R-CO-R' + NADPH.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- TISSUE SPECIFICITY: PRESENT IN LIVER AND KIDNEY.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
FAMILY (SDR).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U07051; AAA77670.1; -
DR EMBL: U09244; AAA82159.1; -
DR PROSITE: P800061; ADH_SHORT; 1.
DR PFAM: PF00106; adh_short; 1.
KW Oxidoreductase; NADP.
FT INIT_MET 0 BY SIMILARITY.
FT NP_BIND 9 NADP (BY SIMILARITY).
FT ACT_SITE 193 BY SIMILARITY.
FT VARIANT 23 A -> N (IN CLONE RCBR6).
FT VARIANT 79 V -> E (IN CLONE RCBR6).
FT VARIANT 91 A -> A (IN CLONE RCBR6).
FT VARIANT 214 H -> Q (IN CLONE RCBR6).
FT SEQUENCE 276 AA; 30321 MW; EAF92E81 CRC32;

DHCA_MOUSE Length: 276 February 14, 2000 08:02 Type: P Check: 1426

```

RC STRAIN-SPRAGUE-DAWLEY; TISSUE-TESTIS;
RA MEDLINE; 95220378.
RT "Cloning and expression of cationic reductase from rat testis."
RL Eur. J. Biochem. 228:473-479(1995).
CC -1- CATALYTIC ACTIVITY: R-COH-R' + NADP(+) -> R-CO-R' + NADPH.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOSOL.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC FAMILY (SDR).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X84349; CAA59088.1; -
DR EMBL; X95886; CAA6230.1; -
DR PROSITE; PS00061; ADH_SHORT; 1.
DR PFAM; PF00106; adh_short; 1.
KW Oxidoreductase; NADP; Acetylation.
FT INITIATION 0 BY SIMILARITY.
FT MOTIFS 1 3 ACETYLATION (BY SIMILARITY).
FT ACT_SITE 193 183 NADP (BY SIMILARITY).
FT ACT_SITE 193 183 BY SIMILARITY.
SQ SEQUENCE 276 AA; 30447 MW; D9BC158B CRC32;
DHCA_RAT Length: 276 February 14, 2000 08:02 Type: P Check: 3202 ..
1 SSDREVALVT GANKGIFAI VDLCKRFGL DYVLTADES RGEHAKQGLQ
51 TEGSLPRHQ LDIDNPOSIR ALRDLQLEY GGLAVLVNNA GLAFKVVDT
101 PFHIAEVTM KTNFGTQDY CKELPIIKP QGRVNVSS VSRALKSS
151 PELQKRFSE TITEELVGL MKKFEDAKK GVHAKGMPN SANGYKIGV
201 TVLSNITVARK LNERREDKI LNAACPGWV RTDAGKRAK KSEGEAEIRP
251 VYLLALPQA BGRHGFQVD KVEPR
11AA_SEQUENCE 1.0
ID DHK1_STRVN STANDARD; PRT; 272 AA.
AC P15342; 1980 (Rel. 15, Created)
DI 01-AUG-1980 (Rel. 15, Last sequence update)
DI 01-AUG-1980 (Rel. 34, Last annotation update)
DI 01-AUG-1980 (Rel. 34, Last annotation update)
DE GRAMICIN POLYKETIDE SYNTHASE PUTATIVE KETOACYL REDUCTASE 1
DE (EC 1.3.1.-) (ORF5).
OS Streptomyces violaceoruber.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomyces.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J11511; CAA77599.1; -
DR EMBL; J11511; CAA77599.1; -
DR PROSITE; PS00061; ADH_SHORT; 1.
DR PFAM; PF00106; adh_short; 1.
DR PFAM; PF00678; adh_short; 2; 1.
KW Antibiotic biosynthesis; Oxidoreductase; NAD.
FT ACT_SITE 157 157 BY SIMILARITY.
FT ACT_SITE 157 157 BY SIMILARITY.
SQ SEQUENCE 261 AA; 27080 MW; 12DF5B3 CRC32;

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X16300; CAA34368.1; -
DR EMBL; X16144; CAA34263.1; -
DR PIR; S05397; S05397.
DR HSSP; P03617; 11PK.
DR PROSITE; PS00061; ADH_SHORT; 1.
DR PFAM; PF00678; adh_short; 1.
DR PFAM; PF00678; adh_short; 2; 1.
KW Antibiotic biosynthesis; Oxidoreductase; NAD.
FT ACT_SITE 168 45 NAD (BY SIMILARITY).
FT ACT_SITE 168 45 BY SIMILARITY.
SQ SEQUENCE 272 AA; 28393 MW; 1C3C79A3 CRC32;
DHK1_STRVN Length: 272 February 14, 2000 08:02 Type: P Check: 3104 ..
1 MTATATATA TPGTAKPVA LVGTATGIG LAIRRLAL GARTFLCARD
51 EERLAQVKE LRGEGEDVDG TVCDVADPAQ IAYVAAYQ RYGVVDILVN
101 NAGSSGGAT AEIADLMID VITINLTSV LMTKEVLNAG GMLAKKRGRI
151 INASTGGRQ GVYHAPVPSA SKHGVGLK AGLELARTG ITVAACPGF
201 VEPMAERVR EHYAGIMQVS EEFTRDITN RYPLGRYVET REVAAWEYL
251 VADDAAVTA QALNVCGGIG NY
11AA_SEQUENCE 1.0
ID DHK1_STRVN STANDARD; PRT; 261 AA.
AC P41177;
DI 01-FEB-1995 (Rel. 31, Created)
DI 01-FEB-1995 (Rel. 31, Last sequence update)
DI 01-OCT-1996 (Rel. 34, Last annotation update)
DE MONENSIN POLYKETIDE SYNTHASE PUTATIVE KETOACYL REDUCTASE 1
DE (ORF5).
OS Streptomyces cinnamonensis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomyces.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z11511; CAA77599.1; -
DR PIR; S25079; S25079.
DR HSSP; P19992; 2HSD.
DR PROSITE; PS00061; ADH_SHORT; 1.
DR PFAM; PF00106; adh_short; 1.
DR PFAM; PF00678; adh_short; 2; 1.
KW Antibiotic biosynthesis; Oxidoreductase; NAD.
FT ACT_SITE 157 157 BY SIMILARITY.
FT ACT_SITE 157 157 BY SIMILARITY.
SQ SEQUENCE 261 AA; 27080 MW; 12DF5B3 CRC32;

DHRR_STRCM Length: 261 February 14, 2000 08:02 Type: P Check: 4473

```

1 MTOSTRVAL VTGATSGIGL ATARLLAOG HLYELGARTE SVIATVAL
51 RNDGLEAEO VLDVBDGASV TAFVQAAYDR YGRIDVLYNN AGRSGGCVTA
101 DLDELMDV IDTNLSVER MTRAVLTGG MTRERGRRI NASTAGKOG
151 VVLGAPYSAS KHVYGFTKA LGNELAPTSI TVNACGYV ETPMAQRYAG
201 GYAAADTTE EAILTKFOAK IPLGRSTPE EVAGLIGYLA SDRAASITGO
251 ALWVCGGLGN F

11AA_SEQUENCE 1.0 STANDARD: PRT: 252 AA.
ID DTE_BACSU
AC P39577:
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE DTE PROTEIN.
GN DTE OR IPA-1R.
OS Bacillus subtilis.
CC Bacteria; Firmicutes; Bacillus/Clostridium group;
CC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-168;
RC MEDLINE; 95020537.
RA GLASER P., KUNESCU F., ARNOLD M., COMDART M.P., GONZALES W.,
RA HULLO M.F., IONESCU M., LUBROCHINSKY B., MARCELINO L., MOSZER I.,
RA PRASEBAN E., SANTANA M., SCHNEIDER E., SCHWEIZER J., VERTES A.,
RA RAPOPORT G., DAVCHIN A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
RT kb region from 325 degrees to 333 degrees."
RT Mol. Microbiol. 10:371-384(1993).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE; 95318144.
RA PERGO M., GLASER P., MINUTELLO A., STRAUCH M.A., LEOPOLD K.,
RA FISCHER W.;
RT "Incorporation of D-alanine into lipoteichoic acid and wall teichoic
RT acid in Bacillus subtilis. Identification of genes and regulation.";
RL J. Biol. Chem. 270:15598-15606(1995).
CC -1- PATHWAY: BIOSYNTHESIS OF D-ALANYL-LIPOTEICHOIC ACID.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC -1- SIMILARITY: TO 3-OXOACYL-LACYL-CARRIER PROTEIN) REDUCTASES.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC FAMILY (SDR).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X73124; CA51557.1;
CC EMBL: Z99123; CAB13880.1;
CC PIR: S39656; S39656.
CC SUBTILIST; BG10547; DTE.
CC PROSITE; PS00061; ADH_SHORT; 1.
CC PFAM; PF00106; adh_short; 1.
CC Oxidoreductase.
CC NP_BIND 9 33 NAD OR NADP (BY SIMILARITY).
CC ACT_SITE 152 152 BY SIMILARITY.
CC FT ACT_SITE 152 152 CC2ALAD CRC32;
CC SEQUENCE 252 AA; 28266 MW; CC2ALAD CRC32;

DTE_BACSU Length: 252 February 14, 2000 08:02 Type: P Check: 2631
1 MKMNTNVLIT TGGSGAGIGLE LAKRLLEIGN EVIIGRSEA KLAKAKOOLP

```

```

51 NHTKOCVYA DRSGREALYE WALKREYPLN VLVNNGIOL EIDFKGTIE
101 LVEDGDEIEL NFOAPYLSA LFTPLMKOP EALIVQVTSQ LAFNPLAVYP
151 VYCATRALH FSLLTNNHY ROTSVEVTEM APPRVDTGLN QKSRDKGLT
201 YRGISSEEV QYFLDKLKEG KQETNERVE GLDAPRADY DLFQOMNTO
251 EN

11AA_SEQUENCE 1.0 STANDARD: PRT: 165 AA.
ID DSB_B_VIBAL
AC Q56578;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE DISULFIDE BOND FORMATION PROTEIN B (DISULFIDE OXIDOREDUCTASE)
DE (FRAGMENT).
GN DSB.
OS Vibrio alginolyticus.
CC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-B138-2;
RC NAKAMURA T., ENOMOTO H., ENOMOTO T.;
RL Submitted (JUL-1996) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: REQUIRED FOR DISULFIDE BOND FORMATION IN SOME
CC PERIPLASMIC PROTEINS SUCH AS PHO OR OMPA. ACTS BY OXIDIZING
CC THE DSB PROTEIN (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D83728; BAA12087.1;
CC DR Oxidoreductase;
CC FT DOMAIN 1 16
CC FT TRANSMEM 3 33 POTENTIAL.
CC FT TRANSMEM 17 51 PERIPLASMIC (POTENTIAL).
CC FT DOMAIN 34 51 POTENTIAL.
CC FT TRANSMEM 52 67 POTENTIAL.
CC FT DOMAIN 68 74 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 75 92 POTENTIAL.
CC FT DOMAIN 93 147 PERIPLASMIC (POTENTIAL).
CC FT TRANSMEM 148 165 POTENTIAL.
CC FT NON_TER 165
CC FT SEQUENCE 165 AA; 18745 MW; PA018063 CRC32;

DSB_B_VIBAL Length: 165 February 14, 2000 08:02 Type: P Check: 2736
1 MTLNLSNLF SKGRSLMILL LFEVFEFC ALYRQHWML APCVMCIYER
51 VYAMGCVAA IVGLMAPNRP IFRWGLIGM GLSKRGILL AOOHVDYQFN
101 PSPATCDLF VTFPSMRPN QNAPWFEAY GDCKRIWQF IDLSMPQWLV
151 VIFAGNLIAL ALIVI

11AA_SEQUENCE 1.0 STANDARD: PRT: 367 AA.
ID DSI_HUMAN
AC P28562;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE DUAL SPECIFICITY PROTEIN PHOSPHATASE 1 (EC 3.1.3.48) (EC 3.1.3.16)
DE (MAP KINASE PHOSPHATASE-1) (MKP-1) (PROTEIN-TYROSINE PHOSPHATASE
DE CL100) (DUAL SPECIFICITY PROTEIN PHOSPHATASE VH1).
DE DUSP1 OR PTPN10 OR MKP1 OR CL100 OR VH1.

```

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-FORESKIN;
 RX MEDLINE: 93024952.
 RA KEYES S.M., EMSLIE E.A.;
 RT "Oxidative stress and heat shock induce a human gene encoding a
 RL protein-tyrosine phosphatase.";
 RL Nature 359:644-647(1992).
 CC -1- FUNCTION: DUAL SPECIFICITY PHOSPHATASE THAT DEPHOSPHORYLATES MAP
 CC KINASE ERK2 ON BOTH THR-183 AND TYR-185.
 CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
 CC PROTEIN TYROSINE + ORTHOPHOSPHATE.
 CC -1- INDUCTION: BY OXIDATIVE STRESS AND HEAT SHOCK
 CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
 CC TYROSINE PHOSPHATASE FAMILY. DUAL SPECIFICITY SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X68277; CAA48338.1; -;
 DR PIR: S29090; S29090.
 DR HSSP: P51452; 1YHR.
 DR MIM: 600714; -;
 DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE: PS50056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE: PS50054; TYR_PHOSPHATASE_DUAL; 1.
 DR PFAM: PF00581; Rhodanese; 1.
 DR PFAM: PF00782; DSPc; 1.
 KW Hydrolyase; Cell cycle;
 FT DOMAIN 175 367 CATALYTIC.
 FT ACT_SITE 258 258 BY SIMILARITY.
 FT SEQUENCE 367 AA; 39297 MW; DA323420 CRC32;
 DUS1_HUMAN Length: 367 February 14, 2000 08:02 Type: P Check: 5401 ..
 1 MYMEVGLTDA GGLRALGER AAQCLLDRC SFAFNAGHI AGSVNREFST
 51 IYRRRAKAM GLEHIVPNAE LRGRLLAGAY HAVVLIDERS AALGAKRDG
 101 TLALAAGALC REARAQVEF LGGYEAFSA SCEPLCSKOS TPKGLSLPLS
 151 TSVPSDAESG CSSCSTPLYD QGGVEILPF LYLGSAYHAS RKMDLDALGI
 201 TALINVSANC PNHEGHYQ KSIPIVEDNHK ADISSWFENEA IDFLDSIKNA
 251 GGRAVEHQA GISRSATICL AYLMRTNRYK LDEAFEVQO RRSITSPNFS
 301 FMGQLLOFES QVLAHCSAE AGSPAMAVLD RGTSTTVEN FVSLPVPVHT
 351 NSALNYLSP ITTSPSC
 //AA_SEQUENCE 1.0
 ID DUS1_MOUSE STANDARD: PRT: 367 AA.
 AC P28563;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE DUAL SPECIFICITY PROTEIN PHOSPHATASE 1 (EC 3.1.3.48) (EC 3.1.3.16)
 DE (MAP KINASE PHOSPHATASE-1) (MPK-1) (PROTEIN-TYROSINE PHOSPHATASE
 DE 3CH134) (PROTEIN-TYROSINE PHOSPHATASE ERP).
 GN DUSP1 OR PTPN10 OR MKP1 OR 3CH134 OR PTPN16.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-BALB/C;
 RX MEDLINE: 92158357.
 RA CHARLES C.H., ABLER A.S., LAU L.F.;
 RT "CDNA sequence of a growth factor-inducible immediate early gene and
 RL characterization of its encoded protein.";
 RL Oncogene 7:187-190(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-LIVER;
 RX MEDLINE: 93360956.
 RA NOGUCHI T., KETZ R., CHEN L., MATTER M.-G., CARRASCO D., BRAVO R.;
 RT "Structure, mapping, and expression of *erp*, a growth factor-inducible
 RT gene encoding a nontransmembrane protein tyrosine phosphatase, and
 RL effect of *erp* on cell growth.";
 RL Mol. Cell. Biol. 13:5195-5205(1993).
 RN [3]
 RP CHARACTERIZATION
 RX MEDLINE: 94037096.
 RA SUN H., CHARLES C.H., LAU L.F., TONKS N.K.;
 RT "MKP-1 (3CH134), an immediate early gene product, is a dual
 RT specificity phosphatase that dephosphorylates MAP kinase in vivo.";
 RL Cell 75:487-493(1993).
 CC -1- FUNCTION: DUAL SPECIFICITY PHOSPHATASE THAT DEPHOSPHORYLATES MAP
 CC KINASE ERK2 ON BOTH THR-183 AND TYR-185.
 CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
 CC PROTEIN TYROSINE + ORTHOPHOSPHATE.
 CC -1- INDUCTION: BY GROWTH FACTORS.
 CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
 CC TYROSINE PHOSPHATASE FAMILY. DUAL SPECIFICITY SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X61940; CAA43944.1; -;
 DR EMBL: S64851; AAB27882.1; -;
 DR PIR: S24411; S24411.
 DR HSSP: P51452; 1YHR.
 DR MGD: MG1:105120; PTPN16.
 DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE: PS50056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE: PS50054; TYR_PHOSPHATASE_DUAL; 1.
 DR PFAM: PF00581; Rhodanese; 1.
 DR PFAM: PF00782; DSPc; 1.
 KW Hydrolyase; Cell cycle;
 FT DOMAIN 175 367 CATALYTIC.
 FT ACT_SITE 258 258
 FT MUTAGEN 258 258 C->S: LOSS OF ACTIVITY.
 FT SEQUENCE 367 AA; 39369 MW; 17B90585 CRC32;
 DUS1_MOUSE Length: 367 February 14, 2000 08:02 Type: P Check: 5871 ..
 1 MYMEVGLTDA GGLRALGER AAQCLLDRC SFAFNAGHI AGSVNREFST
 51 IYRRRAKAM GLEHIVPNAE LRGRLLAGAY HAVVLIDERS ASDGAKRDG
 101 TLALAAGALC REARSQVEF LGGYEAFSA SCEPLCSKOS TPKGLSLPLS
 151 TSVPSDAESG CSSCSTPLYD QGGVEILPF LYLGSAYHAS RKMDLDALGI
 201 TALINVSANC PNHEGHYQ KSIPIVEDNHK ADISSWFENEA IDFLDSIKNA
 251 GGRAVEHQA GISRSATICL AYLMRTNRYK LDEAFEVQO RRSITSPNFS
 301 FMGQLLOFES QVLAHCSAE AGSPAMAVLD RGTSTTVEN FVSLPVPVHT
 351 NSALNYLSP ITTSPSC

```

!!AA_SEQUENCE 1.0 STANDARD; PRT; 367 AA.
ID DUSL_RAT
AC 064623;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE DUAL SPECIFICITY PROTEIN PHOSPHATASE 1 (EC 3.1.3.48) (EC 3.1.3.16)
DE (MAP KINASE PHOSPHATASE-1) (MPK-1) (PROTEIN-TYROSINE PHOSPHATASE
DE CL100).
GN DUSP1 OR CL100.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-LUNG;
RA MUDA M., SCHLEGEL W., ARKINSTALL S.;
RL Submitted (JAN-1995) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: DUAL SPECIFICITY PHOSPHATASE THAT DEPHOSPHORYLATES MAP
CC KINASE ERK ON BOTH THR-183 AND TYR-185.
CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
CC PROTEIN TYROSINE + ORTHOPHOSPHATE.
CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
CC TYROSINE PHOSPHATASE FAMILY. DUAL SPECIFICITY SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X84004; CA55828.1;
CC DR HSSP: P1452; IYR: TYR_PHOSPHATASE_1;
CC DR PROSITE: PS00383; TYR_PHOSPHATASE_2; 1.
CC DR PROSITE: PS00056; TYR_PHOSPHATASE_2; 1.
CC DR PROSITE: PS00054; TYR_PHOSPHATASE_DUAL; 1.
CC DR PFAM: PF00581; Rhodanese; 1.
CC DR PFAM: PF00782; DSpC; 1.
CC KW Hydrolase; Cell cycle.
CC FT DOMAIN 175 367 CATALYTIC.
CC FT ACT_SITE 258 258 BY SIMILARITY.
CC FT SEQUENCE 367 AA; 39541 MW; 9DE26E3 CRC32;
SQ
DUSL_RAT Length: 367 February 14, 2000 08:02 Type: P Check: 6443
1 MMEVEGIIDA GGLRALRER AAQCILLDR SFPAFNAGHI VGSVNVREIT
51 IYVRRAKGM GLEHIVPTE LRGRILAGAY HAVYLLDERS AALDGAKRDG
101 TLALAGALC REARSTQVF LOGGYEAFSA SCPELCSKQS TPNGLSLPIS
151 TSPDSEBSG CSSCSTPLYD QGPFVILSF LYLGSAYHAS RKMIDLALCI
201 TALINVSANC PNHEGHYQY KSIPEVDNRK ADISSWENEA IDEIDSICKA
251 GGRVPHQCA GISRSATICL AYLMRTNRYK IDEAFEFYKQ RRSIIIPNFS
301 FMGOLLPES QVLAHCASAE AGSPAMAVLD RGTSTTVFN EPVSIPIVHT
351 NSALNTYOSP ITTSPSC
!!AA_SEQUENCE 1.0 STANDARD; PRT; 355 AA.
ID DVRL_BRARE
AC P35621;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE DVRL-1 PROTEIN PRECURSOR.
GN VGI OR DVRL-1.
OS Brachydanio rerio (zebrafish) (zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;

```

```

OC Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Cyprinidae; Rasbora; Danio.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE: 94009920.
RA HELDE R.A., GRUNWALT D.J.;
RL The DVRL-1 (vgl) transcript of zebrafish is maternally supplied and
RT distributed throughout the embryo.
RL Dev. Biol. 159:418-426(1993).
CC -1- FUNCTION: SERVES TO FACILITATE THE DIFFERENTIATION OF EITHER
CC MESODERM OR ENDODERM EITHER AS A COFACTOR IN AN INSTRUCTIVE
CC SIGNAL OR BY PROVIDING PERMISSIVE ENVIRONMENT.
CC -1- SUBUNIT: HOMODIMER (PROBABLY).
CC -1- TISSUE SPECIFICITY: ABUNDANT IN OVARIES AND EGGS, AND EQUALLY
CC DISTRIBUTED AMONG ALL BLASTOMERS.
CC -1- DEVELOPMENTAL STAGE: CONCENTRATED IN THE STREAMING CYTOPLASM IN
CC JUST-FERTILIZED EGGS. EVENLY PARTITIONED DURING CLEAVAGE AMONG ALL
CC BLASTOMERS. ABSENT IN THE YOLK CELL DURING CLEAVAGE, BLASTULA AND
CC GASTRULA STAGES. DISTRIBUTED HOMOGENEOUSLY AMONG ALL CELLS OF THE
CC GASTRULA.
CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: U00931; AAC27347.1;
CC DR HSSP: P18075; BMP.
CC DR ZFIN: ZDB-GENE-980526-389; VGL.
CC DR PROSITE: PS00250; TGF-BETA; 1.
CC DR PFAM: PF00019; TGF-beta; 1.
CC KW Growth factor; Mitogen; Glycoprotein; Signal.
CC FT SIGNAL 1 15 POTENTIAL.
CC FT PROPEP 16 240 POTENTIAL.
CC FT CHAIN 241 355 DVRL-1 PROTEIN.
CC FT DISULFID 254 320 BY SIMILARITY.
CC FT DISULFID 283 352 BY SIMILARITY.
CC FT DISULFID 287 354 INTERCHAIN (BY SIMILARITY).
CC FT DISULFID 319 319 POTENTIAL.
CC FT CARBOHYD 108 108 POTENTIAL.
CC FT CARBOHYD 179 179 POTENTIAL.
CC FT CARBOHYD 179 296 POTENTIAL.
CC FT SEQUENCE 355 AA; 40201 MW; B8B2B96C CRC32;
SQ
DVRL_BRARE Length: 355 February 14, 2000 08:02 Type: P Check: 2951
1 MFLYLRAQL LTLISCSPAE DGLVOEKLF LSSMGLMSRP KSHHAAPVS
51 QMKRIFQAS KQVNDPCVY SEYGVNGNIV REMODQSLI SAPAVHSFNC
101 VRKHLFFNNS VLEVEQLSL AOLEMKFKQD LLLGLPHVFS VDLIRVLRKT
151 LKGVTHESR KLQSQTLSP GAHASVLYNL TNLASWRRK ENNEGQDEL
201 QVHNNMLH DHAYQIPDI HATLVVVSIN PLQCSRRKR SASIYLPVP
251 SNVCKPRRTY IDFKDVGWD WILAPOGYLA NYCHGCEPFP LSESLNGTNH
301 AILQTLVHSF DPKGTPOPC VPILKSPISW IYDNDNDVY LRHYEDVAVD
351 ECGCR
!!AA_SEQUENCE 1.0 STANDARD; PRT; 255 AA.
ID E128_DROME
AC P06761;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)

```

DE ECDYSONE-INDUCED PROTEIN 28/29 KD.
GN EIP28/29.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
RN Ephydroidea; Drosophilidae; Drosophila.
RP SEQUENCE FROM N.A.
RC STRAIN-CANTON-S;
RX MEDLINE: 87060956.
RA CHERBAS L., SCHULZ R.A., KOEHLER M.M.D., SAVAKIS C., CHERBAS P.;
RT "Structure of the EIP28/29 gene, an ecdysone-inducible gene from
RT Drosophila.";
RL J. Mol. Biol. 189:617-631(1986).
CC -1- SIMILARITY: TO BACTERIAL AND EUKARYOTIC PEPTIDE METHIONINE
CC SULFOXIDE REDUCTASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X58286; CAA41223.1;
DR EMBL: X04024; CAA27657.1;
DR EMBL: X04521; CAA26205.1;
DR PIR: A24254; A24234.
DR FLYBASE: FBgn0000365; Eip71CD.
DR PFAM: PF01625; PMSR; 1.
KW Alternative splicing.
FT VASPLIC 79
FT CONFLICT 252 252 MISSING (IN ISOFORM EIP29).
FT SEQUENCE 255 AA; 28332 MW; 61E72253 CRC32;
EIP28_DROME Length: 255 February 14, 2000 08:02 Type: P Check: 68
1 MSLLITSSVT HPELKLSTV RNEQKELNIS PVHDVAVTRA TATEMGCFW
51 GAESLXGATR GVLRTVGTA GGSSDLPYR KMGDHEVLE IDYDPTVIST
101 KELLDLFVNN HEYGLTTPK RQVASLLYH DEQKQVANA SKLEQERRA
151 PEITTEIAS KENFPAPAY HOKYRLQGHK DLASSINLSP KILDTSYAT
201 KLNGYLAGVG GIEQKAEAE TTGSDAHPA VLLPRGAER GPGSLLTWP
251 NVARR
11AA-SEQUENCE 1.0
ID ELAC-TRIVU STANDARD; PRT: 102 AA.
AC 029143;
DT 15-JUL-1998 (Rel. 36 Created)
DT 15-JUL-1998 (Rel. 36 Last sequence update)
DT 15-DEC-1998 (Rel. 39 Last annotation update)
DE EARLY LACTATION PROTEIN PRECURSOR.
GN ELP
OS Trichosurus vulpecula (Brush-tailed possum).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Metatheria; Dipodomorpha; Phalangeridae; Trichosurus.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 21-30.
RC TISSUE-MAMMARY GLAND;
RA FIORE C.P., GRIGOR M.R.;
RT "A novel mammary protein expressed by the mammary gland only during
RT the early lactation and related to the Kunitz proteinase inhibitors.";
RL Arch. Biochem. Biophys. 330:59-64(1996).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED ONLY DURING EARLY LACTATION PHASE.
CC -1- TISSUE SPECIFICITY: EXPRESSED BY THE MAMMARY GLAND.

CC -1- PTM: N-GLYCOSYLATED.
CC -1- SIMILARITY: BELONGS TO THE BPTI/KUNITZ FAMILY OF INHIBITORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U34208; AAB08977.1;
DR HSSP: P00974; IPT.
DR PROSITE: PS00280; BPTI_KUNITZ; 1.
DR PFAM: PF00014; Kunitz_BPTI; 1.
KW Serine protease inhibitor; Signal; Lactation.
FT SIGNAL 1 20
FT CHAIN 21 102
FT DISULFD 43 92
FT DISULFD 52 76
FT DISULFD 68 89
FT ACT SITE 53 34
FT CARBOHYD 34 34
FT CARBOHYD 62 62
FT SEQUENCE 102 AA; 11407 MW; D7983676 CRC32;
ELAC-TRIVU Length: 102 February 14, 2000 08:02 Type: P Check: 5524
1 MFTTIALCL ASLVGMSS EKLDRIRAN SLENSRLVP SLCLPSGRC
51 NCDSQLIRYF YNATSHCEY FLYSGCNGNG NNEDSLCCCL KTCRLNKRYN
101 NN
11AA-SEQUENCE 1.0
ID ERP2-YEAST STANDARD; PRT: 215 AA.
AC P39704;
DT 01-FEB-1995 (Rel. 31 Created)
DT 01-FEB-1995 (Rel. 31 Last sequence update)
DT 15-DEC-1998 (Rel. 37 Last annotation update)
DE ERP2 PROTEIN PRECURSOR.
GN ERP2 OR YAL007C OR FUN54.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RX MEDLINE: 95026152;
RA CLARK M.W., KENG T., STORMS R.K., ZHONG W., FORTIN N., ZENG B.,
RA DELANEY S., CHELETTE B.F., BARTON A.B., KABACK D.B., BUSSEY H.;
RT "Sequencing of chromosome I of Saccharomyces cerevisiae: analysis of
RT the 42 kbp SPO7-CENT-CCL15 region.";
RL Yeast 10:535-541(1994).
RN [2]
RP CHARACTERIZATION.
RA MARICHO M., HENTHORN D.C., WILSON R., SOLARI R., THOMAS D.,
RA BERGERON J., ROWLEY A.;
RL Unpublished observations (JUL-1998).
CC -1- FUNCTION: INVOLVED IN VESICULAR PROTEIN TRAFFICKING.
CC -1- SUBUNIT: ASSOCIATES WITH ERP24 AND ERP25.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ENOPLASMIC
CC RETICULUM (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE ERP24/GRP25L FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

```

DR EMBL: L22015; AAC04951.1;
DR PIR: S43448; S43448.
DR PIR: S40896; S40896.
DR SGD: L0004679; ERP2.
DR PFAM: PF01105; EMP24_GP25L; 1.
KW Transport; Protein transport; Transmembrane; Signal;
KW Endoplasmic reticulum.
FT SIGNAL 1 25
FT CHAIN 26 215
FT DOMAIN 26 182
FT TRANSMEM 183 203
FT DOMAIN 204 215
FT SEQUENCE 215 AA; 24063 MW; 50CD1406 CRC32;
ERP2_YEAST Length: 215 February 14, 2000 08:02 Type: P Check: 5386

1 MIKRIALPS FIYILIALY NSVAASSTYA PVAISLAFS KECLYDMVT
51 EDDSLAVGYQ VLTGNFEID EDITAPDSV ITSEKOKYS DELKSEGVG
101 KTFPCSNMY GTALKVEIT LEKEKLTDE HEADVNNNDI IANNAVEED
151 RLNKITITL NYLAREWEN NSTVNSTESR LFWLSILILI IYAVISIAQV
201 LLIOFLFTGR QKNYV

11AA_SEQUENCE 1.0
ID ESI_BRARE STANDARD; PRT: 270 AA.
AC 090237;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE ESI PROTEIN PRECURSOR.
GN ESI.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
OC Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Cyprinidae; Rasbora; Danio.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-EYE;
RX MEDLINE: 97410119.
RA CHANG H., GILBERT W.;
RT "A novel zebrafish gene expressed specifically in the photoreceptor
RT cells of the retina."
RL Biochem. Biophys. Res. Commun. 237:84-89(1997).
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL (POTENTIAL).
CC -1- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN PHOTORECEPTOR CELLS
CC OF THE RETINA.
CC -1- SIMILARITY: BELONGS TO THE ESI FAMILY.
CC This SWISS-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U10403; AAC60261.1;
DR ZFIN: ZDB-GENE-980526-188; ESI.
KW Mitochondrion; Transmembrane; Transmembrane; Signal;
FT TRANSIT 1 270
FT CHAIN 1 270
FT SEQUENCE 270 AA; 30685 MW; 4B69565D CRC32;
ESI_BRARE Length: 270 February 14, 2000 08:02 Type: P Check: 1627

1 MASRALIAK QAAALVROP ACLMHGGDW GNMGNINIAV VESGCGWMDG
51 TDIHEAAYTM YHLSRNGARF QIFAPNOOM HVDHMKROP SSSDRNINIM
101 ESARFSGOG MMOMNDISKL DANSDAVIF PGHGIYKYM STFSKGDKDC

151 KLNDDVERVL KDFHARKPI GLSSNAPILA CRVLPSEVI MGYEDESSR
201 WGRWPTNNY QAVKSMGARH NTRPEEAVY DEKKVISTP TFWMDYHY
251 HYFDGIGN VAKHYWMTAK

11AA_SEQUENCE 1.0
ID ETPA_MYCLE STANDARD; PRT: 318 AA.
AC 033096;
DT 15-DEC-1999 (Rel. 39, Created)
DT 15-DEC-1999 (Rel. 39, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE ELECTRON TRANSFER FLAVOPROTEIN ALPHA-SUBUNIT (ALPHA-ETP) (ELECTRON
DE TRANSFER FLAVOPROTEIN LARGE SUBUNIT) (ETFLS).
ETPA OR FIX OR MCB637.04.
OS Mycobacterium leprae.
GN Bacteria; Firmicutes; Actinobacteria; Mycobacteriaceae;
OC Actinomycetales; Corynebacteriales; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RA OLIVER K., HARRIS D., PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL Submitted (SEP-1997) to the EMBL/Genbank/DBJ databases.
DE FUNCTION: THE ELECTRON TRANSFER FLAVOPROTEIN SERVES AS A SPECIFIC
DE ELECTRON ACCEPTOR FOR OTHER DEHYDROGENASES. IT TRANSFERS THE
DE ELECTRON TO THE MAIN RESPIRATORY CHAIN VIA ETP-OHIDIONE
DE OXYDEREDUCTASE (ETP DEHYDROGENASE) (BY SIMILARITY).
CC -1- COTRANSFER: CONTAINS ONE MOLECULE OF FAD PER DIMER (BY SIMILARITY).
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT.
CC -1- SIMILARITY: BELONGS TO THE ETP ALPHA-SUBUNIT / FIX FAMILY.
CC This SWISS-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: Z99263; CAB16419.1;
DR PROSITE: PS00696; ETP_ALPHA; 1.
DR PFAM: PF00766; ETP_Alpha; 1.
KW Electron transport; Flavoprotein; FAD.
FT NE_BIND 257 285
FT SEQUENCE 318 AA; 31837 MW; 9CF35B35 CRC32;
ETPA_MYCLE Length: 318 February 14, 2000 08:02 Type: P Check: 3764

1 MAELVLEH TEGALKKVA ELITAAVYG EPAAYVGTG GTSAPLYDGL
51 KTAGAKIYV AESDAADXYL ITPVDVLA LAESSAPAV LLAATDKE
101 IGRGLARIG SGLVDVAV REGAVGVAV EGGEIVEAO ANGDPVITV
151 RAGAVEQPA EGAGQVAVE VPAPAEVATK ITPAPAVVD NRPDLTEAV
201 VVSGRGVGS ADNFVEAL ADSIGAAGA SRAADSGYV PGFOFISQTS
251 KTVSPOLYIA LGISGALQHR AGMOTSKIY AVNKDEADI FEIADGVVG
301 DLFVAPOLT DGIKARKG

11AA_SEQUENCE 1.0
ID EXBD_NEIGO STANDARD; PRT: 144 AA.
AC 006434;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE BIOPOLYMER TRANSPORT EXBD PROTEIN.
GN EXBD.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
RN [1]

```



```

DR EMBL: AE000210; AAC74177.1;
DR EMBL: D90745; BAA55901.1;
DR PIR: B42147; B42147.
DR HSSP: P19992; 2HSD.
DR ECOSGENE: EG11318; FABG.
DR PROSITE: PS00061; ADH_SHORT; 1.
DR PFAM: PF00106; adh_short; 1.
DR Fatty acid biosynthesis: Oxidoreductase: NADP.
KM NP_BIND 10 34 NADP (BY SIMILARITY).
FT ACT_SITE 151 151 BY SIMILARITY.
FT CONFLICT 30 30 A -> G (IN REF. 1).
SQ SEQUENCE 244 AA; 25560 MW; 07319E62 CRC32;

FABG_ECOLI Length: 244 February 14, 2000 08:02 Type: P Check: 3492 ..

1 MNEGRALV TGASRGIGRA IAEITLARGA KYIGTATSEN GAQALSDYLG
51 ANGKGLMNV TDRPASTESVL EKIRAEQEV DLYVNNAGIT RDNLNMRKD
101 EKNNDIEN LSVFRLSKA YNRAAMKRR GRIITGSV GTMNGGQAN
151 YAAKAGLIG FSKSLAREVA SRGITVNVVA PGFIETDMR ALSDDQKAGI
201 LAQVPAGRLG GAQETANAVA FLASDEAYI TSETLVHNG MYMV

!!AA_SEQUENCE 1.0
ID FABG_HAEMIN STANDARD: PRT: 242 AA.
AC P43713;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 3-OXOACYL-[ACYL-CARRIER PROTEIN] REDUCTASE (EC 1.1.1.100) (3-KETOACYL-
DE ACYL CARRIER PROTEIN REDUCTASE).
GN FABG OR H10155.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
RN [1]
RP SEQUENCE FROM N.A.
RC SPRATIN-RD / KW20;
RX MEDLINE: 95350630.
RA FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
RA KRELVAGE A.R., BUTT C.J., TOMS J.-F., DOUGHERTY B.A., MERRICK J.M.,
RA KRENNER K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODER A., KELLEY J.M.,
RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOW E., COTTON M.D.,
RA UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SADER D.M., BRANDON R.C.,
RA FINE L.D., FRICHMAN J.L., FUHRMANN J.L., GEORGAGAN N.S.M.,
RA GHEHM C.L., McDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
RA VENTER J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT Influenzae Rd.";
RL Science 269:496-512(1995).
CC -!- CATALYTIC ACTIVITY: (3R)-3-HYDROXYACYL-[ACYL-CARRIER PROTEIN] +
CC NADP(+) = 3-OXOACYL-[ACYL-CARRIER PROTEIN] + NADH.
CC -!- PATHWAY: FIRST REDUCTION STEP IN THE FATTY ACID BIOSYNTHESIS
CC PATHWAY.
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC FAMILY (SDR).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
-----
DR EMBL: U33701; AAC21824.1;
DR HSSP: P19992; 2HSD.
DR TIGR: H10155;
DR PROSITE: PS00061; ADH_SHORT; 1.

```

```

DR PFAM: PF00106; adh_short; 1.
DR PFAM: PF00678; adh_short; 1.
DR Fatty acid biosynthesis: Oxidoreductase: NADP.
KM NP_BIND 8 32 NADP (BY SIMILARITY).
FT ACT_SITE 149 149 BY SIMILARITY.
SQ SEQUENCE 242 AA; 25507 MW; CBFCADF9 CRC32;

FABG_HAEMIN Length: 242 February 14, 2000 08:02 Type: P Check: 4159 ..

1 MOKGIALVTG STGIGRAIA EELSKGAV IGATSEKGA EASAVYLDK
51 GKGLVNVTD KESIEITLBD IKNDGDDI LNVNAGITRD NLNMRKDE
101 WEDIMOTNLT SVYHLKRAMI RSMKRRFR IINIGSVGS TGNPGQTVYC
151 AAKAGVGFSS KSLAKEVAAR GITVNVVAPG FIATDMTEVL TDEKAGILS
201 NVPAGRLGEA KDIAKAVAFI ASDDAGYITG TTIHVNGLY LS

!!AA_SEQUENCE 1.0
ID FABG_PSEAE STANDARD: PRT: 247 AA.
AC O54438;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 3-OXOACYL-[ACYL-CARRIER PROTEIN] REDUCTASE (EC 1.1.1.100) (3-KETOACYL-
DE ACYL CARRIER PROTEIN REDUCTASE).
GN FABG.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC Pseudomonas.
RN [1]
RP SEQUENCE FROM N.A.
RC KUTCHMA A.J., HOANG T.T., SCHWEIZER H.P.;
RA Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RL -!- CATALYTIC ACTIVITY: (3R)-3-HYDROXYACYL-[ACYL-CARRIER PROTEIN] +
CC NADP(+) = 3-OXOACYL-[ACYL-CARRIER PROTEIN] + NADPH.
CC -!- PATHWAY: FIRST REDUCTION STEP IN THE FATTY ACID BIOSYNTHESIS
CC PATHWAY.
CC -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC FAMILY (SDR).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
-----
DR EMBL: U91631; AAB94395.1;
DR PROSITE: PS00061; ADH_SHORT; 1.
DR PFAM: PF00106; adh_short; 1.
DR PFAM: PF00678; adh_short; 1.
KM Fatty acid biosynthesis: Oxidoreductase: NADP.
FT NP_BIND 10 34 NADP (BY SIMILARITY).
FT ACT_SITE 154 154 BY SIMILARITY.
SQ SEQUENCE 247 AA; 25585 MW; C5B95CE0 CRC32;

FABG_PSEAE Length: 247 February 14, 2000 08:02 Type: P Check: 9922 ..

1 MSLOGVNVLT TGASRGIGOA IALEIGRLGA VVIGTATSAS GAETIAETLK
51 ANGVEGAGLV LDVSSDESVA ATLEHIOHL GQPLIVVNA GITRDLNLR
101 MDDEMFQV NTNLSLYRL SKAVLRGMR ARMGRINIG SVYGANGNG
151 QNVYAAKAGV LEFTRLAR EVGSRATVNV AVARGFIDTD MPELPEAQR
201 EALGQIPLG RLGADEIAR VVGFLASDGA AVYGVATVPV NGCMYMS

!!AA_SEQUENCE 1.0

```

ID FABG_VIBHA STANDARD; PRT: 244 AA.
 AC P55336;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE 3-OXOACYL-[ACYL-CARRIER PROTEIN] REDUCTASE (EC 1.1.1.100) (3-KETOACYL-
 DE ACYL CARRIER PROTEIN REDUCTASE).
 GN FABG.
 OS Vibrio harveyi.
 CC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN: 96134997.
 RX MEDLINE: 96134997.
 RA SHEN Z, BYERS D.M.;
 RT "Isolation of Vibrio harveyi acyl carrier protein and the fabg, acpp,
 RL and fadB genes involved in fatty acid biosynthesis.";
 CC CC Bacteriol. 178:571-573(1996).
 CC CC -1- CATALYTIC ACTIVITY: (3R)-3-HYDROXYACYL-[ACYL-CARRIER PROTEIN] +
 CC CC NADP(+) -> 3-OXOACYL-[ACYL-CARRIER PROTEIN] + NADPH.
 CC CC -1- PATHWAY: FIRST REDUCTION STEP IN THE FATTY ACID BIOSYNTHESIS
 CC CC PATHWAY.
 CC CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 CC CC FAMILY (SDR).
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC CC the European Bioinformatics Institute. There are no restrictions on its
 CC CC use by non-profit institutions as long as its content is in no way
 CC CC modified and this statement is not removed. Usage by and for commercial
 CC CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC CC or send an email to license@isb-sib.ch).
 CC CC
 DR EMBL: U39441; AAC43589.1;
 DR HSSP: P19992; 2HSD.
 DR PROSITE: PS00061; ADH_SHORT: 1.
 DR PFAM: PF00106; adh_short: 1.
 DR PFAM: PF00678; adh_short_C2: 1.
 KW Fatty acid biosynthesis; Oxidoreductase; NADP.
 FT NP BIND 10 34 NADP (BY SIMILARITY).
 FT ACT SITE 151 151 BY SIMILARITY.
 FT SEQUENCE 244 AA: 25519 MW; AE81EDC CRC32;
 FABG_VIBHA Length: 244 February 14, 2000 08:02 Type: P Check: 5984 ..
 1 MNECKIALY TGASRGIGRA IAEILLVERGA TVIGTATSEG GAAAISEYLG
 51 ENGKIALNV TDVESIEATL KTIIDECGAI DILVNNAGIT RDNLLMRMD
 101 DEMNDIINTN LPIYRMSKA VLRGAMKKRA GRINVGSSV GTGNAGOTN
 151 YAAKAGVIG FTKSMAREVA SRGVYNTVA PGFIETDWTX ALMDODORAT
 201 LSNVPAGRLG DPREIASAVV FLASPEAVYI TGETLVNGG MYWV
 !!AA_SEQUENCE 1.0
 ID FENR_ECOLI STANDARD; PRT: 247 AA.
 AC P28861; P11007;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE FERREDOXIN-NADP REDUCTASE (EC 1.18.1.2) (FNR) (FLAVODOXIN REDUCTASE)
 DE (FLXR) (FLNR) (METHYL VIIOLOGEN RESISTANCE PROTEIN A) (DRI).
 GN FPR OR MYWV.
 OS Escherichia coli.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 RN (1)
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-25.
 RC STRAIN-K12 / C600;
 RX MEDLINE: 93194782.
 RA BIANCHI V., REICHARD P., ELIASSEN R., PONTIS E., KROOK M.,
 RA JOERNVALL H., HAGGARD-LJUNGQUIST E.;
 RT "Escherichia coli ferredoxin NADP+ reductase: activation of E. coli
 RT anaerobic ribonucleotide reduction, cloning of the gene (fpr), and
 RT overexpression of the protein.";
 RT J. Bacteriol. 175:1590-1595(1993).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE: 93347969
 RA PLUNKETT G., III, BURLAND V.D., DANIELS D.L., BLATTNER F.R.;
 RT "Analysis of the Escherichia coli genome. III. DNA sequence of the
 RT region from 87.2 to 89.2 minutes.";
 RL Nucleic Acids Res. 21:3391-3398(1993).
 RN (3)
 RP PRELIMINARY SEQUENCE OF 1-134 FROM N.A.
 RC STRAIN-K12 / MC4100;
 RX MEDLINE: 93015762.
 RA TRONIGER V., BOOS W., SWEET G.;
 RT "Molecular analysis of the glpFKX regions of Escherichia coli and
 RT Shigella flexneri.";
 RL J. Bacteriol. 174:6981-6991(1992).
 RN (4)
 RP PRELIMINARY SEQUENCE OF 1-128 FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE: 88198002.
 RA MORIMOTO M.;
 RT "Isolation and characterization of methyl viologen-sensitive mutants
 RT of Escherichia coli K-12.";
 RL J. Bacteriol. 170:2136-2142(1988).
 RN (5)
 RP SEQUENCE OF 1-13.
 RX MEDLINE: 95050480.
 RA JENKINS C.M., WATERMAN M.R.;
 RT "Flavodoxin and NADPH-flavodoxin reductase from Escherichia coli
 RT support bovine cytochrome P450C17 hydroxylase activities.";
 RL J. Biol. Chem. 269:27401-27408(1994).
 RN (6)
 RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).
 RX MEDLINE: 97293095.
 RA INEDMAN M., BIANCHI V., EKUND H.;
 RT "The three-dimensional structure of flavodoxin reductase from
 RT Escherichia coli at 1.7-A resolution.";
 RL J. Mol. Biol. 268:147-157(1997).
 CC CC -1- FUNCTION: TOGETHER WITH FLAVODOXIN IS INVOLVED IN THE REDUCTIVE
 CC CC ACTIVATION OF COBALAMIN-INDEPENDENT METHIONINE SYNTHASE, PYRUVATE
 CC CC FORMATE LYASE AND ANAEROBIC RIBONUCLEOTIDE REDUCTASE. ALSO
 CC CC PROTECTS AGAINST SUPEROXIDE RADICALS DUE TO METHYL VIIOLOGEN IN THE
 CC CC PRESENCE OF OXYGEN.
 CC CC -1- CATALYTIC ACTIVITY: REDUCED FERREDOXIN + NADP(+) -> OXIDIZED
 CC CC FERREDOXIN + NADPH.
 CC CC -1- COFACTOR: FAD.
 CC CC -1- SUBUNIT: MONOMER.
 CC CC -1- SIMILARITY: TO OTHER FERREDOXIN NADP REDUCTASES.
 CC CC -1- CAUTION: REF. 3 SEQUENCE WAS INCORRECT DUE TO A FRAMESHIFT.
 CC CC REF. 4 AUTHORS INCORRECTLY ASSIGNED TO BE PART OF FPR, THE C-
 CC CC TERMINAL OF GLPX.
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC CC the European Bioinformatics Institute. There are no restrictions on its
 CC CC use by non-profit institutions as long as its content is in no way
 CC CC modified and this statement is not removed. Usage by and for commercial
 CC CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC CC or send an email to license@isb-sib.ch).
 CC CC
 DR EMBL: L04757; AAA23805.1;
 DR EMBL: L19201; AAB03056.1;
 DR EMBL: A5000467; AAC76906.1;
 DR EMBL: Z11767; NOT ANNOTATED CDS.
 DR EMBL: M19644; AAA24189.1; ALT_SEQ.
 DR PIR: A26225; A26225.
 DR PIR: B45248; B45248.
 DR PIR: A47077; A47077.
 DR PIR: S40867; S40867.

PDB: 1PDR: 17-SEP-97.
 DR ECGENE: EGI1518; FPR.
 DR PFAM: PF00175; oxidored_fad: 1.
 KM Oxidoreductase; Flavoprotein; NADP; FAD; 3D-structure.
 FT INIT_MET 0 0
 FT NP_BIND 49 75 FAD.
 FT NP_BIND 109 125 NADP (RIBOSE PART) (BY SIMILARITY).
 SQ SEQUENCE 247 AA: 27620 MW: 586558A CRC32:
 FENR_ECOLI Length: 247 February 14, 2000 08:02 Type: P Check: 3071 ..
 1 ADVATGKVK VQNTDALS LTVHAPLPF TAGQTKLGL EIDERVORA
 51 YSVNSPDNP DLEFLVTP DGKISPLLA LKPGEVQV SEAGFVLD
 101 EVPRCTIWM LATGTAIGPY LSLIQGLKD DFKNLVLYH AARAADLSY
 151 LPMQELERK YGKLRITQY VSRETAGSL TGRIPALIES GELESTIGLP
 201 MKETSHVWL CGNPQVHDT QQLIKETROM TKHLRRPQH MTAEHYW
 11AA_SEQUENCE 1.0 STANDARD: PRT: 135 AA.
 ID FENR_SHIFL
 AC P28901;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE FERREDOXIN-NADP REDUCTASE (EC 1.18.1.2) (FNR) (FLAVODOXIN REDUCTASE)
 DE (FLXR) (METHYL VIOLOGEN RESISTANCE PROTEIN A) (FRAGMENT).
 GN FPR OR MYRA.
 OS Shigella flexneri.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Shigella.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-M4243;
 RX MEDLINE: 93015762.
 RA TUNINGER V., BOOS W., SWEET G.;
 RT "Molecular analysis of the glpFK regions of Escherichia coli and
 RT Shigella flexneri."
 RL J.Bacteriol. 174:6981-6991(1992).
 CC -1- FUNCTION: TOGETHER WITH FLAVODOXIN IS INVOLVED IN THE REDUCTIVE
 CC ACTIVATION OF COBALAMIN-INDEPENDENT METHIONINE SYNTHASE. PYRUVATE
 CC FORMATE LYASE AND ANAEROBIC RIBONUCLEOTIDE REDUCTASE. ALSO
 CC PROTECTS AGAINST SUPEROXIDE RADICALS DUE TO METHYL VIOLOGEN IN THE
 CC PRESENCE OF OXIGEN (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: REDUCED FERREDOXIN + NADP(+) - OXIDIZED
 CC FERREDOXIN + NADPH.
 CC -1- COFACTOR: FAD.
 CC -1- SIMILARITY: TO OTHER FERREDOXIN NADP REDUCTASES.
 CC -1- CAUTION: THE SEQUENCE SHOWN BELOW WAS CORRECTED TO BE IN AGREEMENT
 CC WITH THE CORRESPONDING E. COLI SEQUENCE.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z11766; CAA77813.1; ALT_SEQ.
 DR PIR: S23906; S23906.
 DR HSSP: P28861; 1PDR.
 KW Oxidoreductase; Flavoprotein; NADP; FAD.
 FT NP_BIND 50 76
 FT NP_BIND 110 126 NADP (RIBOSE PART) (BY SIMILARITY).
 FT NP_BIND 135 135
 FT NON_TER 135 135
 SQ SEQUENCE 135 AA: 14878 MW: F6A7ADSD CRC32:
 FENR_SHIFL Length: 135 February 14, 2000 08:02 Type: P Check: 1251 ..
 1 MADWTKAVT KYQNTWDALE SILVHAPVLP FTAGQTKLGL LEIDERVOR

51 AYSVNSPDNP DLEFLVTP DGKISPLLA LKPGEVQV VSETAGFVLD
 101 DEVPDCTELW LATGTAIGP YLSIQGLKD LDNRK
 11AA_SEQUENCE 1.0 STANDARD: PRT: 203 AA.
 ID FIBR_AKCKO
 AC P28891;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE FIBRINOLASE (EC 3.4.24.72) (FIBRINOLYTIC PROTEINASE).
 DE Agkistrodon contortrix conortrix (Southern copperhead).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Lepidosaurs;
 OC Squamata; Scleroglossa; Serpentes; Colubroidae; Viperidae; Crotalinae;
 CC Agkistrodon.
 RN [1]
 RP SEQUENCE.
 RC TISSUE-VENOM:
 RC MEDLINE: 93278288.
 RX RANDOLPH A.S. JR., MASIARZ F.R.;
 RA MARKLAND F.S. JR., MASIARZ F.R.;
 RT "Amino acid sequence of fibrinolase, a direct-acting fibrinolytic
 RT enzyme from Agkistrodon contortrix conortrix venom."
 RL Protein Sci. 1:350-600(1992).
 CC -1- FUNCTION: THIS PROTEIN IS A ZINC PROTEASE FROM SNAKE VENOM THAT
 CC ACTS IN HEMORRHAGE. IT CLEAVES FIBRINOGEN, A ZINC CHAIN
 CC AND CLEAVAGE OF 413-LYS-1-LEU-414 IN ALPHA CHAIN OF FIBRINOGEN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
 CC METALLOPROTEASE); ALSO KNOWN AS THE REPTOLYSIN SUBFAMILY.
 CC PIR: A41827; HYSNFA.
 DR HSSP: P34179; 2AUG.
 DR PFAM: PF01421; Reptolysin; 1.
 DR PFAM: PF01421; Reptolysin; 1.
 KW Hydrolase; Metalloprotease; Zinc; Venom.
 FT MOD_RES 1 1
 FT VARIANT 1 1
 FT METAL 143 143
 FT ACT_SITE 144 144
 FT METAL 147 147
 FT METAL 153 153
 FT METAL 158 158
 FT DISULFID 158 165
 FT DISULFID 160 182
 FT DISULFID 189 189
 FT VARIANT 192 192
 SQ SEQUENCE 203 AA: 22908 MW: DC49D0E9 CRC32:
 FIBR_AKCKO Length: 203 February 14, 2000 08:02 Type: P Check: 7566 ..
 1 QORFPORYQV LYIADHRAN TKINGDSKI RQWHAIVNT INIYPLNI
 51 QFTVLEIEM SNQDLITVTS VSHDITLASFQ NWRETDILRR QORDNAQLLT
 101 AIDFDDITVG LAYVGMQQL KSTGVYQDH SAINLVALT MAHELGHNNG
 151 MNHNGQCHC GANSCVMAAM LSDQPSKLPF DCSKKDYQTF LTVNPDCLL
 201 NKP
 11AA_SEQUENCE 1.0 STANDARD: PRT: 335 AA.
 ID FIMH_SALTY
 AC P37925;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE FIMH PROTEIN PRECURSOR.
 GN FIMH.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Salmonella.

RN [1]
 RP SEQUENCE FROM N.A.
 RL SWENSON D.L., CLEGG S.;
 CC -1- FUNCTION: INVOLVED IN REGULATION OF LENGTH AND MEDIATION OF
 CC ADHESION OF TYPE 1 FIBRILLAE (BUT NOT NECESSARY FOR THE PRODUCTION
 CC OF FIBRILLAE). BELONGS TO THE FIMH / LPED FAMILY.
 CC -1- SIMILARITY: BELONGS TO THE FIMH / LPED FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L19338; AAA75420.1; -
 DR STYGENE; SG10507; FIMH.
 DR Fimbria; Signal.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 335 FIMH PROTEIN.
 SO SEQUENCE 335 AA; 36056 MW; 2C61967B CRC32;
 FIMH_SALTY Length: 335 February 14, 2000 08:02 Type: P Check: 6074 ..
 1 MKTISALLA GTALFFTHPA LATVCRNSNG TADIFYDLS DVFTSGNNQP
 51 GOVVLPEKS AMGVNATCP AGTVNVTYR SYVELPYQS TEGNFKYKL
 101 NDYLLGAMST TDSVAGVSP PRNYILMGVD YNVSOQKFPQ VODSKLYFKL
 151 KVRPELNMV TIRQTFMTY YVTSIGDAL STPYTISYS GKVEVPQCE
 201 VNAGVVEED FGDIGASLFS QAGAGNRPQ VTPQRTIAI KCTNVAQAY
 251 LSNRLAEKA SGOAVSDNP DLGKVYVANSN GTPLPNNLS SKIPFHDNN
 301 AARVGIRAW PISVIGIRPA EGPFTARGYL RVDYD
 11AA_SEQUENCE 1.0
 ID FLIH_BACSU STANDARD; PRT; 208 AA.
 AC P23449;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE PROBABLE FLAGELLAR ASSEMBLY PROTEIN FLIH.
 GN FLIH.
 OS Bacillus subtilis.
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Bacillus.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN: 68.
 RX MEDLINE: 91258343.
 RA ALBERTINI A.M., CARABO T., CRABB W.D., SCOFFONE F., GALIZZI A.;
 RT The flia locus of Bacillus subtilis is part of a large operon coding
 RT for flagellar structures, motility functions, and an AtPase-like
 RT polypeptide.
 RL J. Bacteriol. 173:3573-3579(1991).
 CC -1- FUNCTION: NEEDED FOR FLAGELLAR REGROWTH AND ASSEMBLY.
 CC -1- SIMILARITY: TO FLIH IN OTHER BACTERIA AND ALSO TO FLIT.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X56049; CA39522.1; -
 DR EMBL: 299112; CAB13496.1; -

DR PIR: S14496; S14496.
 DR PIR: C42365; C42365.
 DR SUBMITLIST: BG10242; FLIH.
 KW Flagella.
 SO SEQUENCE 208 AA; 23788 MW; 448BA68A CRC32;
 FLIH_BACSU Length: 208 February 14, 2000 08:02 Type: P Check: 1150 ..
 1 MARVEEADR ISQANSHE NRRQIEQK NDMAERKRL IEKARGEFE
 51 OGVALGRAEA MKOYALIQ ANTTMSRK AYEDKLEDAV EELVELAVL
 101 AKRYWQKSD DKEAFLLVQ QVINEVEYD DISIYDPY YETIFQORDE
 151 IQQLLYECCR LGIYADEKAO KGTGYETPF GRVDSVDTO LMQLDKLLT
 201 ALKGAEE
 11AA_SEQUENCE 1.0
 ID FOS_AVINK STANDARD; PRT; 322 AA.
 AC P23050;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last annotation update)
 DE P55-V-FOS TRANSFORMING PROTEIN.
 GN V-FOS.
 OS Avian retrovirus NK24.
 CC Viruses; Retroviridae; Avian type C retroviruses.
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 88062820.
 RA NISHIZAWA M., GOTO N., KAMAI S.;
 RT carries the fos gene as the oncogene.
 RT J. Virol. 61:3733-3740(1987).
 CC -1- SIMILARITY: TO OTHER BZIP PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M18041; AAA66158.1; ALT_INIT.
 DR PIR: B34071; TVFV4.
 DR HSSP: P01100; IFOS.
 DR TRANSFAC: T01453; -
 DR PROSITE: PS00036; BZIP_BASIC; 1.
 DR PFAM: PF00170; bzip; 1.
 KW Transforming protein; Nuclear protein; Phosphorylation; DNA-binding.
 FT DNA_BIND 93 147 BASIC MOTIF.
 FT DOMAIN 119 117 LEUCINE-ZIPPER.
 SO SEQUENCE 322 AA; 34333 MW; 7889C981 CRC32;
 FOS_AVINK Length: 322 February 14, 2000 08:02 Type: P Check: 9324 ..
 1 SODFCIDLAV SSANFVPTV AISTSDIOW LVQFTLISSV ABSQNGHPY
 51 GVPAPPPA YSRPAVLKAP GRGQSIGRR GKVEQSPPEE EKKRRIRRR
 101 NKMAAKCKN RRRLDTLQ AETDLEERK SALQAEIANL LKEKLEFEI
 151 LAHRPACMK PEELRFEEL AAATALDGA PSPAAEEAV ALPLMTEAP
 201 AVPEPEPQS GLEKAPED ELLFSAGPRE ASRSVPMDL PGASSFYASD
 251 WEIAGSGS EDEPCTPVY TCTPCSTYT STVFYPERA DAFPCAANAH
 301 RKGSSNEPS SDSISFTLL AL
 11AA_SEQUENCE 1.0

ID FOS_CHICK STANDARD: PRT: 367 AA.
AC P11939;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE P55-C-FOS PROTO-ONCOGENE PROTEIN.
GN FOS.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 88262231.
RA MOEDERS H., JENWEIN T., ADAMKIEWICZ J., MUELLER R.;
RT "Isolation and structural analysis of a biologically active chicken
RT c-fos cDNA: identification of evolutionarily conserved domains in
RT fos protein.";
RL Oncogene 1:377-385(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 88062857.
RA FUJIMURA K.I., ASHIDA K., NISHINA H., IBA H., MIYAJIMA N.,
RA NISHIZAWA M., KAWAI S.;
RT "The chicken c-fos gene: cloning and nucleotide sequence analysis.";
RL J. Virol. 61:4012-4018(1987).
CC - FUNCTION: C-FOS IS A NUCLEAR PHOSPHOPROTEIN WHICH FORM A TIGHT BUT
CC NON-COVALENTLY LINKED COMPLEX WITH THE C-JUN/AP-1 TRANSCRIPTION
CC FACTOR.
CC - SUBCELLULAR LOCATION: NUCLEAR.
CC - INDUCTION: C-FOS EXPRESSION INCREASES RAPIDLY UPON GROWTH FACTOR
CC STIMULATION OR WOUNDING OF CULTURED CELLS.
CC - SIMILARITY: TO OTHER BZIP PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M37000: AAA48670.1; -
DR EMBL: M18043: AA476823.1; -
DR PIR: A28368: TYCHRS.
DR HSSP: P01100: 1FOS.
DR TRANSFAC: T00123; -
DR PROSITE: PS00036; BZIP_BASIC: 1.
DR FRAM: PF00170; bzip; 1.
KM PROTO-oncogene; Nuclear protein; Phosphorylation; DNA-binding.
FT DNAS_BIND 138 159 BASIC MOTIF.
FT DOMAIN 164 192 LEUCINE-ZIPPER.
FT CONFLICT 90 90 N -> D (IN REF. 2).
SQ SEQUENCE 367 AA: 39004 MW: A3FCDACB CRC32;
FOS_CHICK Length: 367 February 14, 2000 08:02 Type: P Check: 7007 ..

ID FOS_HUMAN STANDARD: PRT: 380 AA.
AC P01100;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 21-NOV-1997 (Rel. 35, Last annotation update)
DE P55-C-FOS PROTO-ONCOGENE PROTEIN (GOS7 PROTEIN).
GN FOS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 83221560.
RA VAN STRAATEN F., MULLER R., CURRAN T., VAN BEVEREN C., VERMA I.M.;
RT "Complete nucleotide sequence of a human c-onc gene: deduced amino
RT acid sequence of the human c-fos protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 80:3183-3187(1983).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (3.05 ANGSTROMS) OF 139-198 IN COMPLEX WITH JUN.
RX MEDLINE: 95115802.
RA GLOVER J.N., HARRISON S.C.;
RT "Crystal structure of the heterodimeric bzip transcription factor
RT c-fos-c-jun bound to DNA.";
RL Nature 373:257-261(1995).
CC - FUNCTION: C-FOS IS A NUCLEAR PHOSPHOPROTEIN WHICH FORM A TIGHT BUT
CC NON-COVALENTLY LINKED COMPLEX WITH THE C-JUN/AP-1 TRANSCRIPTION
CC FACTOR.
CC - SUBCELLULAR LOCATION: NUCLEAR.
CC - INDUCTION: C-FOS EXPRESSION INCREASES RAPIDLY UPON GROWTH FACTOR
CC STIMULATION OR WOUNDING OF CULTURED CELLS.
CC - SIMILARITY: TO OTHER BZIP PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: V01512: CAA24756.1; -
DR EMBL: K06550: AAA52471.1; -
DR PIR: A01342: TVHDF1.
DR PDB: 1FOS: 1O-JUL-95.
DR TRANSFAC: T00123; -
DR MIM: 164810; -
DR PROSITE: PS00036; BZIP_BASIC: 1.
DR FRAM: PF00170; bzip; 1.
KM PROTO-oncogene; Nuclear protein; Phosphorylation; DNA-binding;
FT DNAS_BIND 139 160 BASIC MOTIF.
FT DOMAIN 165 193 LEUCINE-ZIPPER.
FT CONFLICT 380 AA: 40695 MW: CDDAEF3C CRC32;
SQ SEQUENCE 380 AA: 40695 MW: CDDAEF3C CRC32;
FOS_HUMAN Length: 380 February 14, 2000 08:02 Type: P Check: 4677 ..

11AA_SEQUENCE 1.0

11AA_SEQUENCE 1.0

ID FTRC_MAIZE STANDARD; PRT; 152 AA.
 AC P41347;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE FERREDOXIN-THIOREDOXIN REDUCTASE, CATALYTIC CHAIN PRECURSOR
 DE (RC 1.18.-.-) (FTR-C) (FERREDOXIN-THIOREDOXIN REDUCTASE SUBUNIT B)
 DE (FTR-B).
 OS FTRC.
 GN Zea mays (Maize).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC eumhyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
 CC Poaceae; Zea.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94002243.
 RA MARC-MARTIN S., SPIELMANN A., STUTZ E., SCHERMAN P.;
 RT Cloning and sequencing of a corn (Zea mays) nuclear gene coding for
 RT the chloroplast specific catalytic subunit of ferredoxin-thioredoxin
 RT reductase.
 RT Biochim. Biophys. Acta 1183:207-209(1993).
 CC -1- FUNCTION: FTR IS A (4FE-4S) PROTEIN PLAYING A CENTRAL ROLE IN THE
 CC FERREDOXIN/THIOREDOXIN REGULATORY CHAIN. IT CONVERTS AN ELECTRON
 CC SIGNAL (PHOTOREDOXED FERREDOXIN) TO A THIOL SIGNAL (REDUCED
 CC THIOREDOXIN) IN THE REGULATION OF ENZYMES BY REDUCTION OF SPECIFIC
 CC DISULFIDE GROUPS. CATALYZES THE LIGHT-DEPENDENT ACTIVATION OF
 CC SEVERAL PHOTOSYNTHETIC ENZYMES.
 CC -1- SUBUNIT: HETERODIMER OF SUBUNIT A (VARIABLE SUBUNIT) AND SUBUNIT
 CC B (CATALYTIC SUBUNIT).
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X73549; ; NOT_ANNOTATED_CDS.
 DR MAIZEDB: 61547; ;
 KW Oxidoreductase; Iron-sulfur; 4Fe-4S; Chloroplast; Transit peptide.
 FT TRANSIT 1 38 CHLOROPLAST (BY SIMILARITY).
 FT CHAIN 1 38 FERREDOXIN-THIOREDOXIN REDUCTASE.
 FT METAL 91 91 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 FT METAL 93 123 REDOX-ACTIVE (BY SIMILARITY).
 FT METAL 110 110 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 FT METAL 112 112 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 FT METAL 121 121 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 SQ SEQUENCE 152 AA; 16740 MW; 25FEE37 CRC32;
 FTRC_MAIZE Length: 152 February 14, 2000 08:02 Type: P Check: 4642 ..
 1 MSTVTTTGG CGGLPVRPLS TATRGPRRC AVRAQAAGAD ASNDKSVEYM
 51 RRFSEYARR SNTFFCADKT VTAIVYKGLA DHRDILGAPL CPCRHDDKA
 101 AEVAGGFVNC PCVPRERKE CHOMLELPDP NDFAGKQDVI SFEIKEANS
 151 KF
 11AA_SEQUENCE 1.0
 ID GDF8_BOVIN STANDARD; PRT; 375 AA.
 AC M01836; 018829;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-DEC-1998 (Rel. 39, Last annotation update)
 DE GROWTH/DIFFERENTIATION FACTOR 8 PRECURSOR (GDF-8) (MYOSTATIN).
 DE GDF8 OR MSTN OR MH.
 OS Bos taurus (Bovine).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;

OC Bovinae; Bos..
 RN [1]
 RP SEQUENCE FROM N.A. AND VARIANT MH TYR-313.
 RC STRAIN-FRIESIAN; TISSUE-MUSCLE, AND EMBRIO;
 RX MEDLINE: 97458167
 RA KAMBAJUR R., SHARMA M., SMITH T.P.L., BASS J.J.;
 RT "Mutations in myostatin (GDF8) in double-muscled Belgian Blue and
 RT Piedmontese cattle";
 RT Genome Res. 7:910-916(1997).
 RN [2]
 RP SEQUENCE FROM N.A. AND VARIANTS MH LEU-94 AND TYR-313.
 RC STRAIN-HOLSTEIN; TISSUE-SKELETAL MUSCLE;
 RX MEDLINE: 98024153.
 RA MCPHERON A.C., LEE S.-J.;
 RT "Double muscling in cattle due to mutations in the myostatin gene";
 RT Proc. Natl. Acad. Sci. U.S.A. 94:12457-12461(1997).
 CC -1- FUNCTION: ACTS SPECIFICALLY AS A NEGATIVE REGULATOR OF SKELETAL
 CC MUSCLE GROWTH.
 CC -1- SUBUNIT: HOMODIMER, DISULFIDE-LINKED (BY SIMILARITY).
 CC -1- DEVELOPMENTAL STAGE: WIDELY EXPRESSED THROUGHOUT DEVELOPMENT.
 CC LOW LEVELS ARE FOUND UP TO DAY 29 EMBRYOS. LEVELS INCREASE FROM
 CC DAY 31 UP UNTIL LATE GESTATION.
 CC -1- TISSUE SPECIFICITY: SPECIFICALLY EXPRESSED IN DEVELOPING AND ADULT
 CC SKELETAL MUSCLE. HIGHEST LEVELS FOUND IN THE HINDLIMB MUSCLES
 CC M. SEMITENDINOSUS AND M. BICEPS FEMORIS; LOW LEVELS IN OTHER
 CC HINDLIMB MUSCLES.
 CC -1- DISEASE: DEFECTS IN GDF8 ARE THE CAUSE OF THE DOUBLE-MUSCLE
 CC PHENOTYPE OR MUSCULAR HYPERTROPHY (MH), AN AUTOSOMAL RECESSIVE
 CC DISEASE FREQUENTLY FOUND IN THE BELGIAN BLUE AND PIEDMONTESE
 CC CATTLE BREDS. THIS DISEASE IS CHARACTERIZED BY AN INCREASED
 CC NUMBER OF MUSCLE FIBERS (HYPERPLASIA), RESULTING IN AN INCREASE IN
 CC MUSCLE MASS OF 20-25%
 CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF019761; AAB81508.1; ;
 DR EMBL: AF019620; AAB86687.1; ;
 DR HSSP: P18075; 1BMF.
 DR PROSITE: P800250; TGF-BETA; 1.
 KW Growth factor; Cytokine; Glycoprotein; Signal; Disease mutation.
 FT SIGNAL 1 266
 FT PROPEP 1 266
 FT CHAIN 1 375
 FT DISULFID 281 340 GROWTH/DIFFERENTIATION FACTOR 8.
 FT DISULFID 309 372 BY SIMILARITY.
 FT DISULFID 313 374 BY SIMILARITY.
 FT DISULFID 339 339 INTERCHAIN (BY SIMILARITY).
 FT CARBOHYD 47 47
 FT CARBOHYD 71 71
 FT CARBOHYD 94 94
 FT VARIANT 94 94
 FT VARIANT 313 313 C -> L (IN MH; PIEDMONTESE BREED).
 FT VARIANT 14 14 T -> M (IN MH; PIEDMONTESE BREED).
 SQ SEQUENCE 375 AA; 42520 MW; F97B649A CRC32;
 GDF8_BOVIN Length: 375 February 14, 2000 08:02 Type: P Check: 9403 ..
 1 MOKYQISVYI YLFTLLVAGP VDLNENSEQ ENVEKEGICN ACIMRENTTS
 51 SRLIAKIQDI LSKRLLETAP NISKDAIRQL LPKAPPLLEL IDQDFORDA
 101 SSDGSLDDO YHARTETVIT NPTESDLILQ VEGAPKCCFE KSSSKIOYAK
 151 LYKAQIMWYL RPYKTPATVE VOILRLIKPM KDGRTYRGIR SKLDMNGGT
 201 GIWQSIDVAT VLOWMLKQPE SMIGIEIKAL DENGDHVAVT FPEPDGGLT


```

251 PELEVKYDT PKRSRDFGL DCDHSTSR CCRYPPLTVDF EAFGDMWII
301 PRRYKANCY GCECEVFLQK YPHTHLVHQ NFRGSGAPCC TPTKMSPLIM
351 LYFNKGEOII YGKIPAMVVD RCGCS

!!AA_SEQUENCE 1.0
ID GDF8_HUMAN STANDARD; PRT; 375 AA.
AC 014793:
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE GROWTH/DIFFERENTIATION FACTOR 8 PRECURSOR (GDF-8) (MYOSTATIN).
GN GDF8 OR MSTN
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE-SKELETAL MUSCLE;
RX MEDLINE; 9602413.
RA MCPHERRON A.C., LEE S.-J.;
RT "Double muscling in cattle due to mutations in the myostatin gene."
RL Proc. Natl. Acad. Sci. U.S.A. 94:12457-12461(1997).
CC -1- FUNCTION: ACTS SPECIFICALLY AS A NEGATIVE REGULATOR OF SKELETAL
CC MUSCLE GROWTH.
CC -1- SUBUNIT: HOMODIMER, DISULFIDE-LINKED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF019627; AAB86694.1; -
DR MIM; 601788; -
DR PROSITE; PS00250; TGF-BETA; 1.
DR PFM; PFO0019; TGF-beta; 1.
KW Growth factor; Cytokine; Glycoprotein; Signal.
FT SIGNAL 1 266
FT PROPEP 2 375
FT CHAIN 267 375
FT DISULFID 281 340
FT DISULFID 309 372
FT DISULFID 313 374
FT DISULFID 339 374
FT CARBOHYD 71 71
SQ SEQUENCE 375 AA; 42750 MW; D0EC1323 CRC32;

GDF8_HUMAN Length: 375 February 14, 2000 08:02 Type: P Check: 1814

```

```

AC 008689:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE GROWTH/DIFFERENTIATION FACTOR 8 PRECURSOR (GDF-8) (MYOSTATIN).
GN GDF8 OR MSTN
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-CD-1; TISSUE-SKELETAL MUSCLE;
RX MEDLINE; 97284412.
RA MCPHERRON A.C., LAWLER A.M., LEE S.-J.;
RT "Regulation of skeletal muscle mass in mice by a new TGF-beta
RT superfamily member."
RL Nature 387:83-90(1997).
CC -1- FUNCTION: ACTS SPECIFICALLY AS A NEGATIVE REGULATOR OF SKELETAL
CC MUSCLE GROWTH.
CC -1- SUBUNIT: HOMODIMER, DISULFIDE-LINKED (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN DEVELOPING AND ADULT
CC SKELETAL MUSCLE. WEAK EXPRESSION IN ADIPOSE TISSUE.
CC -1- DEVELOPMENTAL STAGE: FIRST DETECTED AT DAY 9.5 POST-CONITUM IN
CC ONE-THIRD OF DEVELOPING SOMITES. AT DAY 10.5, EXPRESSED IN THE
CC MYOTOME COMPARTMENT OF SOMITES. AT LATER STAGES OF DEVELOPMENT,
CC DETECTED IN A WIDE RANGE OF DEVELOPING MUSCLES. EXPRESSION
CC CONTINUES IN ADULTHOOD.
CC -----
CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U84005; AAC53167.1; -
DR HSSP; P18075; BMP.
DR MGD; MGI:95691; MSTN.
DR PROSITE; PS00250; TGF-BETA; 1.
DR PFM; PFO0019; TGF-beta; 1.
KW Growth factor; Cytokine; Glycoprotein; Signal.
FT SIGNAL 1 267
FT PROPEP 2 375
FT CHAIN 268 375
FT DISULFID 282 341
FT DISULFID 310 373
FT DISULFID 314 375
FT DISULFID 340 375
FT CARBOHYD 72 72
SQ SEQUENCE 376 AA; 42921 MW; F1ED196 CRC32;

GDF8_MOUSE Length: 376 February 14, 2000 08:02 Type: P Check: 2293

```

AC 018928; 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE GROWTH/DIFFERENTIATION FACTOR 8 PRECURSOR (GDF-8) (MYOSTATIN).
 GN GDF8 OR MSTN.
 OS Papilio hamadryas (Hamadryas baboon).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae;
 OC Papio.
 RP [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE-SKELETAL MUSCLE;
 RX MEDLINE: 98024153.
 RA MCPHERSON A.C., LEE S.-J.;
 RT "Double muscling in cattle due to mutations in the myostatin gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 94:12457-12461(1997).
 CC -1- FUNCTION: ACTS SPECIFICALLY AS A NEGATIVE REGULATOR OF SKELETAL
 CC MUSCLE GROWTH.
 CC -1- SUBUNIT: HOMODIMER, DISULFIDE-LINKED (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: AF019619; AAB86686.1;
 DR PROSITE: PS00250; TGF-BETA; 1.
 DR PFAM: PF00019; TGF-beta; 1.
 KW Growth factor; Cytokine; Glycoprotein; Signal.
 FT SIGNAL 1 ?
 FT PROPEP 267 266 POTENTIAL.
 FT CHAIN 267 375 GROWTH/DIFFERENTIATION FACTOR 8.
 FT DISULFID 281 340 BY SIMILARITY.
 FT DISULFID 309 372 BY SIMILARITY.
 FT DISULFID 313 374 BY SIMILARITY.
 FT DISULFID 339 339 INTERCHAIN (BY SIMILARITY).
 FT CARBOHYD 71 71 POTENTIAL.
 SQ SEQUENCE 375 AA; 42688 MW; 61C8D2B7 CRC32;
 GDF8_PAPA Length: 375 February 14, 2000 08:02 Type: P Check: 1463 ..
 1 MOKLQIYVI YLPMIIVAGP VDLNENSEQK ENVEKEGLCN ACTWRONTKS
 51 SRLFAIKIQI LSKRLLETAP NISKDAIRQL LPRAPPLREL IDQYDVQRDD
 101 SSDGSEDDD YHATTEIIT MPTESDLMO VDGKPKCFE KFSKRIQYNK
 151 YKQAQLWIYL RPYKPTIYF VOILRLIKPM KDGRTYTGIR SLKIDMNGT
 201 GIMOSIDVKT VLQNLKQPE SNLGEIKAL DENGHDIAVT PPGEGEDGLN
 251 PLEVKYVDT PKRSRDFGL DCDHSTESR CCRYPITVDF EALGWDWIIA
 301 PKRYKANVCS GECEYFLOK YPHTHLVHOA NPGSAGPC TPTKMSINM
 351 LYFNGKEQII YKRIPIAVYD RCGCS
 11AA_SEQUENCE 1.0
 ID GDF8_PIG STANDARD: PRT: 375 AA.
 AC 018931;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE GROWTH/DIFFERENTIATION FACTOR 8 PRECURSOR (GDF-8) (MYOSTATIN).
 GN GDF8 OR MSTN.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.
 RP [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE-SKELETAL MUSCLE;

RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-SKELETAL MUSCLE;
 RX MEDLINE: 98024153.
 RA MCPHERSON A.C., LEE S.-J.;
 RT "Double muscling in cattle due to mutations in the myostatin gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 94:12457-12461(1997).
 CC -1- FUNCTION: ACTS SPECIFICALLY AS A NEGATIVE REGULATOR OF SKELETAL
 CC MUSCLE GROWTH.
 CC -1- SUBUNIT: HOMODIMER, DISULFIDE-LINKED (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: AF019623; AAB86690.1;
 DR PROSITE: PS00385; AAC8035.1;
 DR EMBL: AF03798; AAC8248.1;
 DR PROSITE: PS00250; TGF-BETA; 1.
 DR PFAM: PF00019; TGF-beta; 1.
 KW Growth factor; Cytokine; Glycoprotein; Signal.
 FT SIGNAL 1 ?
 FT PROPEP 267 266 POTENTIAL.
 FT CHAIN 267 375 GROWTH/DIFFERENTIATION FACTOR 8.
 FT DISULFID 281 340 BY SIMILARITY.
 FT DISULFID 309 372 BY SIMILARITY.
 FT DISULFID 313 374 BY SIMILARITY.
 FT DISULFID 339 339 INTERCHAIN (BY SIMILARITY).
 FT CARBOHYD 71 71 POTENTIAL.
 SQ SEQUENCE 375 AA; 42791 MW; 65CB82B8 CRC32;
 GDF8_PIG Length: 375 February 14, 2000 08:02 Type: P Check: 1805 ..
 1 MOKLQIYVI YLPMIIVAGP VDLNENSEQK ENVEKEGLCN ACMMRONTKS
 51 SRLFAIKIQI LSKRLLETAP NISKDAIRQL LPRAPPLREL IDQYDVQRDD
 101 SSDGSEDDD YHATTEIIT MPTESDLMO VEGKPKCFE KFSKRIQYNK
 151 YKQAQLWIYL RPYKPTIYF VOILRLIKPM KDGRTYTGIR SLKIDMNGT
 201 GIMOSIDVKT VLQNLKQPE SNLGEIKAL DENGHDIAVT PPGEGEDGLN
 251 PLEVKYVDT PKRSRDFGL DCDHSTESR CCRYPITVDF EAFGWDWIIA
 301 PKRYKANVCS GECEYFLOK YPHTHLVHOA NPGSAGPC TPTKMSINM
 351 LYFNGKEQII YKRIPIAVYD RCGCS
 11AA_SEQUENCE 1.0
 ID GDF8_RAT STANDARD: PRT: 376 AA.
 AC 035312;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE GROWTH/DIFFERENTIATION FACTOR 8 PRECURSOR (GDF-8) (MYOSTATIN).
 GN GDF8 OR MSTN.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.
 RP [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE-SKELETAL MUSCLE;

RX MEDLINE: 98024153.
 RA MCPHERSON A.C., LEE S.-J.;
 RT "Double muscling in cattle due to mutations in the myostatin gene."
 RL Proc. Natl. Acad. Sci. U.S.A. 94:12457-12461(1997).
 CC -1- FUNCTION: ACTS SPECIFICALLY AS A NEGATIVE REGULATOR OF SKELETAL
 CC MUSCLE GROWTH.
 CC -1- SUBUNIT: HOMODIMER, DISULFIDE-LINKED (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF019624; AAB86691.1;
 DR PROSITE: PS00250; TGF-BETA: 1.
 DR PFAM: PF00019; TGF-beta: 1.
 KW Growth factor; Cytokine; Glycoprotein; Signal.
 FT SIGNAL 1 267
 FT PROPEP 1 267
 FT CHAIN 268 376
 FT DISULFID 282 341
 FT DISULFID 310 373
 FT DISULFID 314 375
 FT DISULFID 340 340
 FT CARBOHYD 72 72
 SQ SEQUENCE 376 AA; 42829 MW; BFC16D96 CRC32; POTENTIAL.
 GDF8_RAT Length: 376 February 14, 2000 08:02 Type: P Check: 1496
 1 MOKQDMVYV YLFLVLIAG PVDLNESEK EANYKEKGLC MACAMQNQR
 51 YSRLEAIKIQ ILSKRLLETA PNISKDAIRQ LLPRAPPLRE LIDQYDVRD
 101 SSSGSLDD DYHATTEIIL TPTESDFLM QADGPKCCF KFKSKIQYN
 151 KYVAKQLWIY LRAVKTPTVY FVQILRLIKP MKDGRYTG I RSLKLMSPG
 201 TGIMQSIDVK TVLQNMWKOP ESNLGEIKR LDENGHLAV TFGPGEDGL
 251 NFPLEKVIDT TPKRSRDFG LDCDEHSTES RCRRLPLVD FEAFGDWII
 301 AKRRYANVC GCECEPFLQ KYPHILVHQ ANPRGSAGPC CPTKMSPIN
 351 LYFNKEKQI IYKIPANVY DRGCS
 11AA_SEQUENCE 1.0
 ID GDF8_SHEEP STANDARD: PRT: 375 AA.
 AC 018830;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE GROWTH/DIFFERENTIATION FACTOR 8 PRECURSOR (GDF-8) (MYOSTATIN).
 GN GDF8 OR MSTN.
 OS Ovis aries (Sheep).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
 OC Caprinae; Ovis.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE-SKELETAL MUSCLE;
 RX MEDLINE: 98024153.
 RA MCPHERSON A.C., LEE S.-J.;
 RT "Double muscling in cattle due to mutations in the myostatin gene."
 RL Proc. Natl. Acad. Sci. U.S.A. 94:12457-12461(1997).
 CC -1- FUNCTION: ACTS SPECIFICALLY AS A NEGATIVE REGULATOR OF SKELETAL
 CC MUSCLE GROWTH.
 CC -1- SUBUNIT: HOMODIMER, DISULFIDE-LINKED (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF019624; AAB86691.1;
 DR PROSITE: PS00250; TGF-BETA: 1.
 DR PFAM: PF00019; TGF-beta: 1.
 KW Growth factor; Cytokine; Glycoprotein; Signal.
 FT SIGNAL 1 266
 FT PROPEP 1 266
 FT CHAIN 267 375
 FT DISULFID 281 340
 FT DISULFID 309 372
 FT DISULFID 313 374
 FT DISULFID 339 339
 FT CARBOHYD 48 48
 FT CARBOHYD 71 71
 SQ SEQUENCE 375 AA; 42827 MW; B36084EE CRC32; POTENTIAL.
 GDF8_SHEEP Length: 375 February 14, 2000 08:02 Type: P Check: 1548
 1 MOKQITVYI YLFLMLVAGP VDLNENSEQ ENVEKKGLCN ACLMRQNNKS
 51 SRLPAITQI LSKRLLETA PNISKDAIRQ LLPRAPPLRE LIDQYDVRD
 101 SSSGSLDD DYHATTEIIL TPTESDLAE VQEKPKCCF KFKSKIQHNK
 151 KYVAKQLWIY LRAVKTPTVY FVQILRLIKP MKDGRYTG I RSLKLMSPG
 201 TGIMQSIDVK TVLQNMWKOP ESNLGEIKR LDENGHLAV TFGPGEDGL
 251 NFPLEKVIDT TPKRSRDFG LDCDEHSTES RCRRLPLVD FEAFGDWII
 301 AKRRYANVC GCECEPFLQ KYPHILVHQ ANPRGSAGPC CPTKMSPIN
 351 LYFNKEKQI IYKIPANVY DRGCS
 11AA_SEQUENCE 1.0
 ID GDF8_PANTR STANDARD: PRT: 123 AA.
 AC 028914;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE GLYCOPHORIN B PRECURSOR.
 GN GYPB OR GPB.
 OS Pan troglodytes (Chimpanzee).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homnidae; Pan.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC MEDLINE: 96029146.
 RA HUANG C.H., XIE S.S., SOCHA W., BLUMENFELD O.O.;
 RT "Sequence diversification and exon inactivation in the glycophorin A
 RT gene family from chimpanzee to human."
 RL J. Mol. Evol. 41:478-486(1995).
 CC -1- FUNCTION: THIS PROTEIN IS A MINOR SIALOGLYCOPROTEIN IN ERYTHROCYTE
 CC MEMBRANES.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE GLYCOPHORIN A FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: S79726; AAB35339.1;

DR HSPP: P02724; 1MSR.
 DR PROSITE; PS00312; GLYCOPHORIN A; 1.
 DR PFAM; PF01102; Glycophorin A; 1.
 KW Erythrocyte; Transmembrane; Sialic acid; Glycoprotein; signal.
 FT SIGNAL 1 19 BY SIMILARITY.
 FT CHAIN 20 123 GLYCOPHORIN B.
 FT TRANSMEM 93 113 POTENTIAL.
 SO SEQUENCE 123 AA; 13246 MW; 28850255 CRC32;

GPDA_PANTR Length: 123 February 14, 2000 08:02 Type: P Check: 149

1 MGKRIIFVL LSEIVSISAS STEEVAMHS TSSSVTSYI SSOTNDKMG
 51 DTPATLGAH EVSEISTYV YPEEDNGEM VQVHPSPRP APVYILLIL
 101 CVMAGVIGTI LLISYGRLL IKA

!!AA_SEQUENCE 1.0 STANDARD; PRT; 339 AA.

AC P37606;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE GLYCEROL-3-PHOSPHATE DEHYDROGENASE [NAD+] (EC 1.1.1.8).
 GN GP5A.

OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.

RC STRAIN-K12 / MG1655;
 RA MEDLINE; 94316500.

RT SORIA H.J., BURLAND V., DANIELS D.L., PLUNKETT G. III, BLATTNER F.R.;
 "Analysis of the Escherichia coli genome. V. DNA sequence of the
 region from 76.0 to 81.5 minutes."

RL Nucleic Acids Res. 22:2576-2586(1994).
 CC -1- CATALYTIC ACTIVITY: SN-GLYCEROL-3-PHOSPHATE + NAD(+) -
 GLYCERONE PHOSPHATE + NADH
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE)
 CC -1- SIMILARITY: BELONGS TO THE NAD-DEPENDENT GLYCEROL-3-PHOSPHATE
 DEHYDROGENASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL; U00039; CAB34639.1; -
 DR EMBL; AE000439; AAC76632.1; -
 DR ECGENE; EG20091; GP5A.
 DR PROSITE; PS00957; NAD_G3PDH; 1.
 DR PFAM; PF01210; NAD_GLY3P_dh; 1.
 KW Oxidoreductase; NAD.
 SO SEQUENCE 339 AA; 36361 MW; 085E0E12 CRC32;

GPDA_ECOLI Length: 339 February 14, 2000 08:02 Type: P Check: 4443

1 MNQNAAMTV IGAGSYGTAL AITLARGHE VLMGDPPEH TATLERDNC
 51 AAFIPDPFP DTLHESDIA TALAASRNIL VVPSHVGGE VLRQIKPLMR
 101 PDAFLVATK GLEAETGRLL QDVAREALGD QIPLAVISGD TPAKELANGL
 151 PTAISLASTD QTFADLQQL LHGGSREFRV SNPDFGYOL GQAVKNVIAI
 201 GAGNSDIFG GANARTALIT RGLAEMSRIG AALGADPAFE WGMAGLGDLY
 251 LITCNDOSRN RRFGMGLGOG MDVOSAOEKI GQVVEGYRNT KEVELAHRE
 301 GVEMPTTEEI YQVLYCGRNA REAALTILGR ARKDERSSH

!!AA_SEQUENCE 1.0 STANDARD; PRT; 355 AA.

AC GPRD_HUMAN
 ID P49238;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE PROBABLE G-PROTEIN-COUPLED RECEPTOR GPR13 (V28) (BETA CHEMOKINE
 RECEPTOR-LIKE 1) (CMK-BRL-1) (CMKBLR1).
 GN GPR13.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.

RA RAPPORT C.J., SCHWEICKART V.L., EDDY R.L. JR., SHOWS T.B., GRAY P.W.;
 "The orphan G-protein-coupled receptor-encoding gene v28 is closely
 related to genes for chemokine receptors and is expressed in lymphoid
 and neural tissues."

RL Gene 163:295-299(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 95374679.

RA COMBADIÈRE C., AHUJA S.K., MURPHY P.M.;
 "Cloning, chromosomal localization, and RNA expression of a human
 beta chemokine receptor-like gene."

RT DNA Cell Biol. 14:673-680(1995).
 CC -1- FUNCTION: ORPHAN RECEPTOR. COULD BE A CHEMOKINE RECEPTOR.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN LYMPHOID AND NEURAL TISSUES.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL; U20350; AAA91783.1; -
 DR EMBL; U28934; AAA87032.1; -
 DR GCRDB; GCR_1992; -
 DR GCRDB; GCR_2014; -
 DR MIM; 601470; -
 DR PROSITE; PS00237; G-PROTEIN_RECEPTOR; 1.
 DR PFAM; PF00001; 7tm_1; 1.

KW G-protein coupled receptor; Transmembrane.
 FT DOMAIN 1 31
 FT TRANSMEM 1 31
 FT DOMAIN 32 59
 FT TRANSMEM 32 59
 FT DOMAIN 60 69
 FT TRANSMEM 60 69
 FT DOMAIN 70 90
 FT TRANSMEM 70 90
 FT DOMAIN 91 103
 FT TRANSMEM 91 103
 FT DOMAIN 104 125
 FT TRANSMEM 104 125
 FT DOMAIN 126 142
 FT TRANSMEM 126 142
 FT DOMAIN 143 167
 FT TRANSMEM 143 167
 FT DOMAIN 168 195
 FT TRANSMEM 168 195
 FT DOMAIN 196 215
 FT TRANSMEM 196 215
 FT DOMAIN 216 231
 FT TRANSMEM 216 231
 FT DOMAIN 232 256
 FT TRANSMEM 232 256
 FT DOMAIN 257 273
 FT TRANSMEM 257 273
 FT DOMAIN 274 297
 FT TRANSMEM 274 297
 FT DISULFID 298 355
 FT DISULFID 102 175
 SO SEQUENCE 355 AA; 40396 MW; E809E31D CRC32;

GPDR_HUMAN Length: 355 February 14, 2000 08:02 Type: P Check: 2863

1 MDQPESTVE NFEYDLAEA CYIGDIYVFG TVFSLIYSV IFAIGLVGL
 51 LVVFALTNK KPKSVTDIYL LNLALSDLLF VATLPFTVHY LNEGLHNA

101 MCKFTTAFPF IGFGSGIFFI TVISIDRYLA IYLAANSNN RTVOHGVTIS
 151 LGWMAAILV APOPFMTKQ KENECLGDY EYLOEIMPV LNVEINFLGF
 201 LEPILLMSYC YFRIIQTLES CKNRKKAKAI KLILIVIVF FLFWTPYNYW
 251 IFETIKLYD FFPSCDMRKD LRLALSYTE VAFSHCCINP LIYAFAGEKF
 301 RRYLYHLYGK CLAVLCGRSV HYDFSSSESO RSRHGSVLSS NLYHTSDGD
 351 ALLLL

11AA_SEQUENCE 1.0 STANDARD: PRT: 354 AA.
 ID GPRD_RAT
 AC P35411;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE PROBABLE G PROTEIN-COUPLED RECEPTOR GPR13 (RBS11).
 GN GPR13.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE-SPINAL CORD;
 RX MEDLINE: 94323113.
 RA HARRISON J.R., BARBER C.M., LYNCH K.R.;
 RT "cDNA cloning of a G-protein-coupled receptor expressed in rat spinal
 cord and brain related to chemokine receptors.";
 RL Neurosci. Lett. 169:85-89(1994).
 CC -1- FUNCTION: ORPHAN RECEPTOR. COULD BE A CHEMOKINE RECEPTOR.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: MOST ABUNDANT IN ADULT SPINAL CORD, BRAIN,
 CC KIDNEY, GUT, UTERUS AND TESTES.
 CC -1- PFM: THIS PROTEIN IS NOT N-GLYCOSYLATED WHICH IS UNUSUAL FOR
 CC G-PROTEIN-COUPLED RECEPTORS.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U04808; AAB87093.1; -;
 DR GCRDB: GCR_0268; -;
 DR PROSITE: PS00237; G-PROTEIN_RECEPTOR: 1.
 DR PFM: PFM0001; 7tm_1; 1.
 KW G-protein coupled receptor; Transmembrane.
 FT DOMAIN 1 32 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 33 60
 FT DOMAIN 61 70 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 71 91
 FT DOMAIN 92 104 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 105 126
 FT DOMAIN 127 143 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 144 168
 FT DOMAIN 169 196 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 197 216
 FT DOMAIN 217 232 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 233 257
 FT DOMAIN 258 274
 FT TRANSMEM 275 298
 FT DOMAIN 299 354
 FT TRANSMEM 354 40327 MW: 798DA277 CIRC32;
 FT DISULFID 103 176 BY SIMILARITY
 FT SEQUENCE 354 AA: 40327 MW: 798DA277 CIRC32;
 GPRD_RAT Length: 354 February 14, 2000 08:02 Type: P Check: 4478
 1 MPTSPBELDL ENFEYDSDAE ACYLGIDIVAF GTIFLSTIFS LYTFELGVGN

51 LLYVALTNS RRSKSTIDY LNLIALSDL EVATLPFWTH YLISHESLHN
 101 AMCKLTITAF FIFGFGIPE ITVISIDRYL AYLANSNN RTVOHGVTI
 151 SLGWMAAIL VASPOFMFK RKNECLGDY PEYLOEIMPV LRNSEVNLG
 201 FVLELLMSF CYFRIYRTLF SKNRKKARA IRLILVVV FLFWTPYNI
 251 VIFETIKFY NFPPSGMKR DLRLALSYE TVAFSHCCIN PFIYAFGEK
 301 FRIYRLILYN KCLAVLCGRP VHAGSTIESQ RSRHDSILSS LHYTSGEG
 351 SLLL

11AA_SEQUENCE 1.0 STANDARD: PRT: 247 AA.
 ID GRAB_MOUSE
 AC P04187;
 DT 20-MAR-1987 (Rel. 04, Created)
 DT 20-MAR-1987 (Rel. 04, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE GRANZYME B(G,H) PRECURSOR (EC 3.4.21.79) (CYTOTOXIC CELL PROTEASE 1)
 DE (CCP1) (CTLA-1) (FRAGMENTIN 2).
 GN GZMB OR CTLA-1 OR CTLA-1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE: 86208120.
 RA LOBE C.G., FINLAY B.B., PARANCHYCH W., PATRAU V.H., BLACKLEY R.C.;
 RT "Novel serine proteases encoded by two cytotoxic T
 RT lymphocyte specific genes.";
 RL Science 232:858-861(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC MEDLINE: 89062424.
 RA LOBE C.G., UPTON C., DUGGAN B., EHMAN N., LETELLIER M., BELL J.,
 RA MCFADDEN G., BLACKLEY R.C.;
 RT "Organization of two genes encoding cytotoxic T lymphocyte-specific
 RT serine proteases CCP1 and CCP11.";
 RL Biochemistry 27:6941-6946(1988).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC MEDLINE: 86284960.
 RA BRUNET J.F., DOSSETO M., DENIZOT F., MATTEI M.-G., CLARK W.R.,
 RA HAOUI T.M., FERRIER P., NABHOLZ M., SCHMITT-VERHULST A.M.,
 RA LUCIANI M.F., GOLSTEIN P.;
 RT "The inducible cytotoxic T-lymphocyte-associated gene transcript
 RT CTLA-1 sequence and gene localization to mouse chromosome 14.";
 RL Nature 322:268-271(1986).
 RN [4]
 RP SEQUENCE OF 237-247 FROM N.A.
 RC STRAIN-C57BL/6J;
 RX MEDLINE: 94319082.
 RA KO M.S., WANG X., HORTON J.H., HAGEN M.D., TAKAHASHI N.,
 RA VAEZAKI Y., NADDEV J.H.;
 RT "Genetic mapping of 40 cDNA clones on the mouse genome by PCR.";
 RL Mamm. Genome 5:349-353(1994).
 RN [5]
 RP SEQUENCE OF 21-40.
 RX MEDLINE: 87215932.
 RA MASSON D., TSCHOPE J.;
 RT "A family of serine esterases in lytic granules of cytolytic T
 RT lymphocytes.";
 RL Cell 49:679-685(1987).
 RN [6]
 RP 3D-STRUCTURE MODELING.
 RX MEDLINE: 89184501.
 RA MURPHY M.E.P., MOULT J., BLACKLEY R.C., GERSHENFELD H.,
 RA WEISSMAN I.L., JAMES M.N.G.;
 RT "Comparative molecular model building of two serine proteinases from
 RT cytotoxic T lymphocytes.";

RL Proteins 4:190-204(1988).
 CC -1- FUNCTION: THIS ENZYME IS NECESSARY FOR TARGET CELL LYSIS IN CELL-MEDIED IMMUNE RESPONSES. IT CLEAVES AFTER ASP. SEEMS TO BE LINKED TO AN ACTIVATION CASCADE OF CASPASES (ASPARTATE-SPECIFIC CYSTEINE PROTEASES) RESPONSIBLE FOR APOPTOSIS EXECUTION. CLEAVES CASPASE-3, -7, -9 AND 10 TO GIVE RISE TO ACTIVE ENZYMES MEDIATING APOPTOSIS (BY SIMILARITY).
 CC CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ASP-|-XAA >> ASP-|-XAA
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC GRANULES OF CYTOLYTIC T-LYMPHOCYTES AND NATURAL KILLER CELLS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1. ALSO KNOWN AS THE TRYPSIN FAMILY. STRONGEST TO OTHER GRANZYMS AND TO MAST CELL PROTEASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X04072; CAA27715.1; -
 DR EMBL: M12302; AAA37383.1; -
 DR EMBL: M22526; AAB61756.1; -
 DR EMBL: U05707; AAB60470.1; -
 DR PIR: A00956; PRMSCL.
 DR PIR: A28952; A28952.
 DR PIR: B26944; B26944.
 DR PDB: 2CP1; 15-OCT-94.
 DR MGD: MGI:109267; GZMB.
 DR PROSITE: PS00134; TRYPSIN_HIS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR PFAM: PF00089; trypsin; 1.
 DR KW Hydrolyase; Serine protease; Zymogen; Signal; T-cell; Cytolysis;
 KM Apoptosis; 3D-structure.
 FT SIGNAL 1 18
 FT PROPEP 19 20 ACTIVATION PEPTIDE.
 FT CHAIN 21 247 GRANZYME B(G/H).
 FT ACT_SITE 64 64 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 108 108 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 203 203 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 49 63 BY SIMILARITY.
 FT DISULFID 142 209 BY SIMILARITY.
 FT DISULFID 173 188 BY SIMILARITY.
 FT CARBOHYD 71 71 POTENTIAL.
 FT CARBOHYD 182 182 POTENTIAL.
 SQ SEQUENCE 247 AA; 27470 MW; E05916CB CRC32;
 GRAB_MOUSE Length: 247 February 14, 2000 08:02 Type: P Check: 92 ..
 1 MKILLILLTL SLASRTAGE IIGGHEVKNP SRPYMALSI KDOQPEALICG
 51 GFLREDEVVL TAAHCEGSI NVTLGHNKIK EQEKTQOYIP MYKCIHPDY
 101 NKRTSNDIM LKTKSKAKR TRAVRPLNP RRNVNVPD VCVVAGWGM
 151 APMGRYSNTL QEVETLVQKD RECESYFKNR YNKINQICAG DKPKTKASR
 201 GDSGPVLCK KVAAGISYSG YKDGSPRAF TKVSSFLSWI KTKMMS
 !!AA_SEQUENCE 1.0 STANDARD: PRT: 267 AA.
 ID HDNA_CLOSO
 AC P50200.
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE 7-ALPHA-HYDROXYSTEROID DEHYDROGENASE (EC 1.1.1.159) (BILE ACID
 DE 7-DEHYDROXYLASE) (?-ALPHA-HSDH).
 OS Clostridium sordeilii.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 CC Clostridium.

FN [1]
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 1-18.
 RC STRAIN-ATCC 9714.
 RX MEDLINE: 94327451.
 RA COLEMAN J.P., HUDSON L.L., ADAMS M.J.;
 RT "Characterization and regulation of the NAD-linked 7 alpha-hydroxysteroid dehydrogenase gene from Clostridium sordeilii";
 RL J. Bacteriol. 166:4863-4874(1994).
 CC -1- CATALYTIC ACTIVITY: 3-ALPHA,7-ALPHA,12-ALPHA-TRIHYDROXY-5-BETA-CHOLANATE + NADP(+) = 3-ALPHA,12-ALPHA-DIHYDROXY-7-OXO-5-BETA-CHOLANATE + NADPH.
 CC -1- SUBUNIT: HOMOTETRAMER (PROBABLE).
 CC -1- INDUCTION: BY BILE ACID. EXPRESSED DURING GROWTH-PHASE.
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES FAMILY (SDR).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: L12058; AAA35556.1; -
 DR HSSP: P25529; 1AHH.
 DR PROSITE: PS00061; ADH_SHORT; 1.
 DR PFAM: PF00106; adh_short; 1.
 DR PFAM: PF00678; adh_short; 2.
 DR OXIDOREDUCTASE: NADP; Bile acid catabolism.
 FT NP_BIND 10 34 NADP (BY SIMILARITY).
 FT ACT_SITE 158 158 BY SIMILARITY.
 SQ SEQUENCE 267 AA; 29177 MW; FCCD9D5A CRC32;
 HDNA_CLOSO Length: 267 February 14, 2000 08:02 Type: P Check: 1047 ..
 1 MNKLEKVAL VTSATRGIGL ASAIKLAONG AIVYGVRRLL EATQELCDKY
 51 KEKGLIKPV FEDAVNIDYI KEMIDITLNN ESKIDILVNN FGTRPREKDL
 101 DLYNGEDDTF FELFNVYNGS VYRLSKLIIP HMIENKGGSI VNISVSGSI
 151 PDISRIIGYGV SKSGVNNITK QIAIQYAKYG IKNVNLVGL IATDAMNSM
 201 PDEFRKSFIS HVPENRIGNP EDIANSVLFE VPSEDSYIT GSILEVSGGY
 251 NLGTPQVAF VQSKVYE
 !!AA_SEQUENCE 1.0 STANDARD: PRT: 296 AA.
 ID HG2A_HUMAN
 AC P04233; Q29832.
 DT 20-MAR-1987 (Rel. 04, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 05-DEC-1999 (Rel. 39, Last annotation update)
 DE HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, GAMMA CHAIN (HLA-DR ANTIGENS
 DE ASSOCIATED IN VARIANT CHAIN) (P33) (CD74 ANTIGEN).
 GN CD74 OR DR4G.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 NC [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE: 86093681.
 RA KUDO J., CHAO L.Y., NARNI F., SAUNDERS G.F.;
 RT "Structure of the human gene encoding the invariant gamma-chain of class II histocompatibility antigens";
 RL Nucleic Acids Res. 13:8827-8841(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 84207945.
 RA STRUBIN M., MACH B., LONG E.O.;
 RT "The complete sequence of the mRNA for the HLA-DR-associated invariant chain reveals a polypeptide with an unusual transmembrane

RT POLARITY.";
 RL EMBL J. 3:869-872(1984).
 [3]
 RP SEQUENCE OF 27-96 FROM N.A.
 RX MEDLINE: 84170234.
 RA CLAESSENS L., LARHAMMAR D., RASK L., PETERSON P.A.;
 RT "CDN clone for the human invariant gamma chain of class II
 RT histocompatibility antigens and its implications for the protein
 RT structure." Acad. Sci. U.S.A. 80:7395-7399(1983).
 RL Proc. Natl. Acad. Sci. U.S.A. 80:7395-7399(1983).
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 86233451.
 RA O'SULLIVAN D.M., LARHAMMAR D., WILSON M.C., PETERSON P.A.,
 RA OVARANTA V.;
 RT "Structure of the human Ia-associated invariant (gamma)-chain gene:
 RT identification of 5' sequences shared with major histocompatibility
 RT complex class II genes." J. S.A. 83:4484-4488(1986).
 RL Proc. Natl. Acad. Sci. U.S.A. 83:4484-4488(1986).
 [5]
 RP X-RAY CRYSTALLOGRAPHY (2:75 ANGSTROMS) OF 103-117.
 RX MEDLINE: 96085023.
 RA GHOSH P., AMAYA M., MELLINS E., WILEY D.C.;
 RT "The structure of an intermediate in class II MHC maturation: CLIP
 RT bound to HLA-DR3." Nature 378:457-462(1995).
 RL Nature 378:457-462(1995).
 RN [6]
 RP STRUCTURE BY NMR OF 134-208.
 RX MEDLINE: 99059718.
 RA JASANOFF A., WAGNER G., WILEY D.C.;
 RT "Structure of a trimetric domain of the MHC class II-associated
 RT chaperonin and targeting protein II." EMBO J. 17:6812-6818(1998).
 RL EMBO J. 17:6812-6818(1998).
 CC -1- FUNCTION: PLAYS A CRITICAL ROLE IN MHC CLASS II ANTIGEN PROCESSING
 CC BY STABILIZING PEPTIDE-FREE CLASS II ALPHA/BETA HETERODIMERS IN A
 CC COMPLEX SOON AFTER THEIR SYNTHESIS AND DIRECTING TRANSPORT OF THE
 CC COMPLEX FROM THE ENDOPLASMIC RETICULUM TO COMPARTMENTS WHERE
 CC PEPTIDE LOADING OF CLASS II TAKES PLACE.
 CC -1- SUBUNIT: NONAMER COMPOSED OF THREE ALPHA/BETA/GAMMA HETEROTRIMERS.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: CONTAINS 1 THYROGLOBULIN TYPE-I DOMAIN.
 CC -1- DATABASE: NAME-PROT: NOTE-CD guide CD74 entry;
 CC WWW: http://www.ncbi.nlm.nih.gov/prot/CD/CD74.htm".
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: X03339; CAA27046.1; -
 CC EMBL: X03340; CAA27047.1; -
 CC EMBL: X03341; AAA36304.1; -
 CC EMBL: X00497; CAA25192.1; -
 CC EMBL: X00497; CAA25193.1; -
 CC PIR: A30060; HLHUG.
 CC PDB: 1A6A; 27-MAY-98.
 CC PDB: 1IE; 15-FEB-99.
 CC KIM: 142790; -
 CC PROSITE: PS00484; THYROGLOBULIN_1; 1.
 CC PRAM: PF00086; thyroglobulin_1; 1.
 CC KW Glycoprotein; MHC; Proteoglycan; Transmembrane; Alternative splicing;
 CC Signal-anchor; 3D-structure.
 FT DOMAIN 1 46 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 47 72 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
 FT DOMAIN 73 296 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 103 117 CLIP.
 FT CARBOHYD 130 130
 FT CARBOHYD 136 136
 FT CARBOHYD 282 282
 FT DOMAIN 228 271 GLYCOSAMINOGLYCAN.
 FT THYROGLOBULIN TYPE I.

FT VARSPLIC 209 272 MISSING (IN SHORT ISOFORM).
 FT CONFLICT 167 T -> R (IN REF. 2 AND 3).
 SO SEQUENCE 296 AA: 33460 MW: 488A044A CRC32:
 HG2A_HUMAN Length: 296 February 14, 2000 08:02 Type: P Check: 9949 ..
 1 MHRRSRSCR EDKRPYMDQ RDLSINNEQ PMLGRPGAP ESKSGRGLY
 51 TGSILVTL LAGQATYAF LYQOQGRDLX LVTSTQNLQ ENLRMKLPK
 101 PKVSKKRRMA TPLMQLPM GALPQGMQN AKKYNMTEQ HYVHLQNLAD
 151 PLKYPLPLKG SEPENITLHK NTMETIMKY FESMHHML FEMSRHSLEQ
 201 KPLDAPKVL TKQGEVSHI PAVHGSFEP KCDENGNILP LQCGSLGIC
 251 WCPFNGTEV PNTSRGHHN CSESLLEDP SSGLYTKOD LGPYPM
 11AA_SEQUENCE 1.0 STANDARD; PRT; 359 AA.
 ID HIS8_SALTY
 AC P10369;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)
 DE HISTIDINOL-PHOSPHATE AMINOTRANSFERASE (EC 2.6.1.9) (IMIDAZOLE ACETOL-
 DE PHOSPHATE TRANSAMINASE).
 GN HIS8.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LT2;
 RX MEDLINE: 89094829.
 RA CAROMONGO M.S., CHARTIOTI L., ALIFANO P., NAPPO A.G., BRUNI C.B.;
 RT "Structure and function of the salmonella typhimurium and Escherichia
 RT coli K-12 histidine operons." J. Mol. Biol. 203:585-606(1988).
 RL J. Mol. Biol. 203:585-606(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LT2;
 RA BARNES W.M., HUSON R.N., WHITTIER R.;
 RL Submitted (Aug-1989) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: L-HISTIDINOL-PHOSPHATE + 2-OXOGLUTARATE =
 CC 3-(IMIDAZOL-4-YL)-2-OXOPROPATE + GLUTAMATE.
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
 CC -1- PATHWAY: EIGHTH STEP IN HISTIDINE BIOSYNTHETIC PATHWAY.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SIMILARITY: BELONGS TO CLASS-II OF PYRIDOXAL-PHOSPHATE-DEPENDENT
 CC AMINOTRANSFERASES.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: X13464; CAA31824.1; -
 CC EMBL: J01804; AAA88616.1; -
 CC PIR: J01584; XNEBHC.
 CC STYGENE: SGI0159; HIS8.
 CC PROSITE: PS00599; AA-TRANSFER CLASS_2; 1.
 CC PRAM: PF00222; aminotran_2; 1.
 CC KW Histidine biosynthesis; Transferase; Aminotransferase;
 CC Pyridoxal phosphate.
 FT BINDING 216 217 PYRIDOXAL PHOSPHATE (PROBABLE).
 FT CONFLICT 148 164 DGTKVFVCSNPTGQ -> TQKWCSEVAPILPBN
 FT CONFLICT 226 226 C -> R (IN REF. 1).
 FT CONFLICT 260 260 L -> S (IN REF. 1).
 FT CONFLICT 264 265 GI -> ES (IN REF. 1).

3D-STRUCTURE MODELING OF 136-196.
 RA MEDLINE: 94259190.
 RT "A homology-based molecular model of the proline-rich homeodomain
 RT protein Prh, from hematopoietic cells."
 RT FEBS Lett. 345:93-98(1994).
 CC -1- FUNCTION: RECOGNIZES THE DNA SEQUENCE 5'-ATTAA-3'. MAY PLAY A ROLE
 CC IN HEMATOPOIETIC DIFFERENTIATION.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
 CC -1- TISSUE SPECIFICITY: LIVER AND PROMYELOCYTIC LEUKEMIA CELL
 CC LINE HL60.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING HEMATOPOIESIS.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X67235; CAA47661.1; -;
 CC EMBL: L16499; AA02988.1; -;
 CC EMBL: Z21533; CAA79730.1; -;
 CC PIR: S26799; S26799.
 CC HSSP: P23441; 1FTT.
 CC TRANSFAC: T02092; -;
 CC PROSITE: PS00027; HOMEBOX_1; 1.
 CC PROSITE: PS50071; HOMEBOX_2; 1.
 CC PIR: PF00046; Homeobox; 1.
 CC Homeobox; DNA-binding; Developmental protein; Nuclear protein.
 CC DOMAIN 1 133 PRO-RICH.
 CC FT DNA_BIND 137 196 HOMEBOX.
 CC FT CONFLICT 115 115 L -> V (IN REF. 2).
 CC FT SEQUENCE 270 AA; 30021 MW; 453E9E6A CRC32;
 HMHP_HUMAN Length: 270 February 14, 2000 08:02 Type: P Check: 3785 ..
 1 MGFHPGPA GAVGVPLA PTLQPAHP PFYIDILG GPAAPPTP
 51 LSPNSSTF LSPYRTPV EPTPIHAPS HNSAALAA YPGGFGGGL
 101 YEPFRVND YTHLRHDP LKPLWSPF LQPLHKRG GVFNSDQT
 151 ELEKKEIQ KLSPEKRL AKMLQSER QKTFQNRRA KMRLLQEN
 201 OSNKKKEELS LDSCDOROD LPSEONKGA LDSCSCSPS ASQEDLESEI
 251 SEDSDQVDI EGDKYFNAG
 11AA_SEQUENCE 1.0
 ID HMHP_MOUSE STANDARD: PRT; 271 AA.
 AC P43120;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HOMEBOX PROTEIN PRH (HOMEBOX PROTEIN HEX).
 GN PRHX OR HEX.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN: C57BL/6J; TISSUE: EMBRYO;
 RX MEDLINE: 93219088.
 RA BEDFORD F.K. ASHWORTH A., ENVER T., WIEDEMANN L.M.;
 RA "Hex: a novel homeobox gene expressed during haematopoiesis and
 RT conserved between mouse and human."
 RT Nucleic Acids Res 21:1245-1249(1993).
 CC -1- FUNCTION: RECOGNIZES THE DNA SEQUENCE 5'-ATTAA-3'. MAY PLAY A ROLE
 CC IN HEMATOPOIETIC DIFFERENTIATION.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING HEMATOPOIESIS.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Z21524; CAA79729.1; -;
 CC HSSP: P23441; 1FTT.
 CC MGD: MGI:96086; HEX.
 CC PROSITE: PS00027; HOMEBOX_1; 1.
 CC PROSITE: PS50071; HOMEBOX_2; 1.
 CC PIR: PF00046; homeobox; 1.
 CC Homeobox; DNA-binding; Developmental protein; Nuclear protein.
 CC DOMAIN 1 134 PRO-RICH.
 CC FT DNA_BIND 138 197 HOMEBOX.
 CC FT SEQUENCE 271 AA; 29966 MW; 96744BE3 CRC32;
 HMHP_MOUSE Length: 271 February 14, 2000 08:02 Type: P Check: 8616 ..
 1 MGFHPGPA GAVGVPLA PTLQPAHP PFYIDILG GPAAPPTP
 51 TLSPNSSTF LSPYRTPV EPTPIHAPS SHHAAALAA AYSGFGGCP
 101 YEPFRVND YTHLRHDP LKPLWSPF LQPLHKRG GVFNSDQT
 151 VELEKKEIQ KLSPEKRL AKMLQSER QKTFQNRRA KMRLLQEN
 201 POSNKKDALD SLDTSCQGO DLPSEONKGA LDSCSCSPS PASQEDDSE
 251 IESDQVDI EGDKYFNAG
 11AA_SEQUENCE 1.0
 ID HPRK_BACSU STANDARD: PRT; 309 AA.
 AC O34483;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE HPR(SER) KINASE/PHOSPHATASE (EC 2.7.1.-) (EC 3.1.3.-).
 GN PTK.
 OS Bacillus subtilis.
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC LAZAREVIC V., SORDO B., RIVOLTA C., REYNOLDS S., MAUEL C.,
 RA KARAMATA D.;
 RA submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP CHARACTERIZATION, AND SEQUENCE OF 1-10.
 RC MEDLINE: 98230327.
 RA REITZER J., HOISCHEN C., TITTEMEYER F., RIVOLTA C., RABUS R.,
 RA STELKE J., KARAMATA D., SAIER M.H. JR., HILLEN W.;
 RA "A novel protein kinase that controls carbon catabolite repression in
 RT bacteria."
 RT Mol. Microbiol. 27:1157-1169(1998).
 RN [3]
 RP CHARACTERIZATION OF THE PHOSPHATASE ACTIVITY.
 RC MEDLINE: 99141583.
 RA KRAVANJA M., ENGELMANN R., DOSSONNET V., BLUGGEL M., MEYER H.E.,
 RA FRANK R., GALINIER A., DEUTSCHER J., SCHNELL N., HENGSTENBERG W.;
 RA "The hpr gene of Enterococcus faecalis encodes a novel bifunctional
 RT enzyme: the hpr kinase/phosphatase."
 RT Mol. Microbiol. 31:59-66(1999).
 CC -1- FUNCTION: THIS KINASE/PHOSPHATASE REGULATES CARBOHYDRATE UPTAKE
 CC AND METABOLISM BY PHOSPHORYLATING/DEPHOSPHORYLATING SER-45 OF THE
 CC PHOSPHOTRANSFERASE SYSTEM (PTS) IN RESPONSE TO
 CC CYTOPLASMIC METABOLITE LEVELS AS A DEVICE FOR THE MODULATION OF
 CC CARBON CATABOLITE REPRESSION.
 CC -1- ENZYME REGULATION: ALLOSTERICALLY ACTIVATED BY METABOLITES SUCH AS

CC FRUCTOSE-1,6-BISPHOSPHATE AND INHIBITED BY INORGANIC PHOSPHATE.
CC -1- MASS SPECTROMETRY: MW=34529; METHOD-MALDI.
CC -1- SIMILARITY: BELONGS TO THE PISK FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF017113; AAC67286.1; -
CC DR EMBL: 299121; CAB15505.1; -
CC DR SUBMITTER: BG14125; PISK.
CC KW Multifunctional enzyme; transferase; Serine/threonine-protein kinase;
CC KM ATP-binding; Hydrolyase.
CC FT INIT MET 147 0
CC NP_BIND 152 154 ATP (POTENTIAL).
CC FT NP_BIND 152 159 ATP (POTENTIAL).
CC SQ SEQUENCE 309 AA; 34571 MW; 3272ECOE CRC32;
HPK_BACSV Length: 309 February 14, 2000 08:02 Type: P Check: 4393
1 ANKRDVME QPNELISGE EGINRPTMS DLSRPGIEIA GYFTYPRER
51 VOLLGTELS FEQLPEEEK KORMDSLCTD VTPAIIISRD MPIDQLIDA
101 SERNGVPLVR SPLKTRLLS RLTFPLESRL APPTAIGHVL VDIYGVLLI
151 TGRSGVXSE TALELVKRGH RLVAADCVET ROEDQDTLVG NAELEHLL
201 EIRGIGIIV MTEFGAGAVR SNKRITIVN LEIWEQGRQY DRGLSEETM
251 KIIDEITKL TIVPRGRNL AVIEVAAM FRLRMGLNA AEGFTKMLD
301 VIEDEGOEE
11AA_SEQUENCE 1.0
ID HUPK_RHLIV STANDARD; PRT; 370 AA.
AC P28153;
DT 01-JUL-1993 (Rel. 26, Created)
DI 01-FEB-1996 (Rel. 33, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYDROGENASE EXPRESSION/FORMATION PROTEIN HUPK.
GN HUPK.
OS Rhizobium leguminosarum (biovar victoriae).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
RA STRAIN-128C3;
RC MEDLINE: 93108466.
RX KEY L., HIDALGO E., PALACIOS J.M., RUIZ-ARGUESO T.;
RT "Nucleotide sequence and organization of an H2-uptake gene cluster
RT from Rhizobium leguminosarum bv. viciae containing a rubredoxin-like
RT gene and four additional open reading frames.";
RL J. Mol. Biol. 228:998-1002(1992).
RN [12]
RP SEQUENCE FROM N.A.
RX MEDLINE: 95020662.
RA IMPERIAL J., KEY L., RUIZ-ARGUESO T.;
RT "HUPK, a hydrogenase-ancillary protein from Rhizobium leguminosarum,
RT shares structural motifs with the large subunit of NIFE hydrogenases
RT and could be a scaffolding protein for hydrogenase metal cofactor
RT assembly.";
RL Mol. Microbiol. 9:1305-1306(1993).
RN [13]
RP SEQUENCE FROM N.A.
RX STRAIN-B10;
RA SCHMITT H., KOKOTER W., THUERING H., KERL V., BAUER T., FUCHS D.,
RA TICHY H., LOTZ W.;
RT Submitted (SEP-1994) to the EMBL/Genbank/DBJ databases.

CC -1- SIMILARITY: BELONGS TO THE HUPK FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X52974; CAA37158.1; -
CC DR EMBL: 236981; CAA85440.1; -
CC DR CONFLICT 240 240 A -> R (IN REF. 1).
CC SQ SEQUENCE 370 AA; 38751 MW; 3F9E2409 CRC32;
HUPK_RHLIV Length: 370 February 14, 2000 08:02 Type: P Check: 1548
1 MTEFGAGTI GIDVTYSRAL ACSVAVKANR PGLTRMEVG ROPEANPVLA
51 GOVESLGFPA OSVAARLAVL AAADIAWNE ERLGASGLL AERLEFLRA
101 LILOMPTEPE ERFADAGRH LRELAASLA ITSHKAGRT SRRLAAAE
151 RLSSAATLIG IPRGDTPLP ETACALILND VEDDHVFAGR RPPPLTISD
201 AEVVARLDE AGYASLPHLS GRIAETGAYA RSASGLPEA PLARLRAR
251 IGDVRLSLQ LTLARTGDF DCASLACGP TPGAGGYGAV ECARGRLHQ
301 IEIGSGGRLA AYRIAPTEM NEHPAPPEVE TLSSPVGAD EAAVRSISRL
351 AYLPDPCVAF EIVNREADA
11AA_SEQUENCE 1.0
ID HYCD_ECOLI STANDARD; PRT; 307 AA.
AC B16430;
DT 01-AUG-1990 (Rel. 15, Created)
DI 01-AUG-1990 (Rel. 15, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE FORMATE HYDROGENLYASE SUBUNIT 4 (FHL SUBUNIT 4) (HYDROGENASE 3
DE COMPONENT D).
GN HYCD OR HEVD.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RA STRAIN-K12 / MC4100;
RC MEDLINE: 90251163.
RX BOEHM R., SAUTER M., BOECK A.;
RT "Nucleotide sequence and expression of an operon in Escherichia coli
RT coding for formate hydrogenlyase components.";
RL Mol. Microbiol. 4:231-243(1990).
RN [12]
RP SEQUENCE FROM N.A.
RX STRAIN-K12 / MC1655;
RA BLATTNER F.R., BLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [1]
RP -1- PATHWAY: HYDROGEN METABOLISM; FHL PATHWAY.
CC -1- SUBUNIT: FHL COMPRISES OF A FORMATE DEHYDROGENASE, UNIDENTIFIED
CC ELECTRON CARRIERS AND A HYDROGENASE (ISOMERASE 3). IN THIS NON-
CC ENERGY CONSERVING PATHWAY MOLECULAR HYDROGEN AND CARBODIOXIDE
CC FROM FORMATE ARE RELEASED.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC -1- PTM: THE DISULFIDE BONDS ARE ESSENTIAL FOR THE INHIBITOR ACTIVITY.
 CC -1- SIMILARITY: BELONGS TO THE CEREAL TRYPSIN/ALPHA-AMYLASE INHIBITOR
 CC FAMILY.
 DR PIR: A01322; WMTA.
 DR HSSP: P01085; 1HSS.
 DR PROSITE: PS00426; CEREAL_TRYP_AMYL_INH. 1.
 DR PFAM: PF00234; TRYF_alpha_aml; 1.
 KW Alpha-amylase inhibitor; Multigene family; Seed.
 FT DISULFID 21 42
 FT DISULFID 29 82
 FT DISULFID 43 98
 FT DISULFID 56 113
 FT VARIANT 65 65
 FT VARIANT 67 67
 FT VARIANT 98 99
 FT CONFLICT 118 118 A -> C (IN REF. 1).
 SO SEQUENCE 123 AA; 13326 MW; 82F56DD7 CRC32;
 IAA2.WHEAT Length: 123 February 14, 2000 08:02 Type: P Check: 4825
 1 SGPWSCNCPA TGYKVSALTG CRAWKLOCV GSQVPEAVLR DCCQGLADIN
 51 NEMCRGDLG SMURSVQEL GVREKENVLP GCRREVKMLT AASVEVCV
 101 PIPNSGDRA GVCYGDWAAV PDV
 IIAA-SEQUENCE 1.0 STANDARD; PRT; 124 AA.
 AC P01084;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE ALPHA-AMYLASE INHIBITOR 0.53.
 OS Triticum aestivum (wheat).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
 CC Poaceae; Triticum.
 RN [1]
 RP SEQUENCE.
 RX MEDLINE: 83127436.
 RA MADDA K., HASE T., MATSUBARA H.;
 RT "Complete amino acid sequence of an alpha-amylase inhibitor in wheat
 RT kernel.";
 RL Biochim. Biophys. Acta 743:52-57(1983).
 RN [2]
 RP REVISIONS TO 119-124.
 RX MEDLINE: 85175148.
 RA MADDA K., WAKABAYASHI S., MATSUBARA H.;
 RT "Complete amino acid sequence of an alpha-amylase inhibitor in wheat
 RT kernel (0.19-inhibitor).";
 RL Biochim. Biophys. Acta 828:213-221(1985).
 RN [3]
 RP DISULFIDE BONDS.
 RX MEDLINE: 84061717.
 RA MADDA K., WAKABAYASHI S., MATSUBARA H.;
 RT "Disulfide bridges in an alpha amylase inhibitor from wheat kernel.";
 RL J. Biochem. 94:863-870(1983).
 CC -1- FUNCTION: ALPHA-AMYLASE INHIBITOR.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- TISSUE SPECIFICITY: ENDOSPERM.
 CC -1- PTM: THE DISULFIDE BONDS ARE ESSENTIAL FOR THE INHIBITOR ACTIVITY.
 CC -1- SIMILARITY: BELONGS TO THE CEREAL TRYPSIN/ALPHA-AMYLASE INHIBITOR
 CC FAMILY.
 DR PIR: A01323; WMTAS.
 DR HSSP: P01085; 1HSS.
 DR PROSITE: PS00426; CEREAL_TRYP_AMYL_INH. 1.
 DR PFAM: PF00234; TRYF_alpha_aml; 1.
 KW Alpha-amylase inhibitor; Multigene family; Seed.
 FT DISULFID 6 115
 FT DISULFID 20 41 OR 20-42.
 FT DISULFID 28 83
 FT DISULFID 42 99 OR 41-99.

SO SEQUENCE 124 AA; 13185 MW; 3FE87339 CRC32;
 IAA5.WHEAT Length: 124 February 14, 2000 08:02 Type: P Check: 4073
 1 SGPWMCYPCG AFQVPALPGC RPLKLCQNG SQVPEAVLRD CCQGLADISE
 51 WPRGALYSM LDMKKEHGV SEQAGTGAF PGCRREVKL TASTAVCR
 101 LPTVNASGD GAVCKDVYA YPDA
 IIAA-SEQUENCE 1.0 STANDARD; PRT; 312 AA.
 AC P55215; 035398;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE CASPASE-2 PRECURSOR (EC 3.4.22.-) (CASP-2) (ICH-1 PROTEASE)
 DE (FRAGMENT).
 GN CASP2 OR ICH1.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 RN [1]
 RP SEQUENCE OF 1-257 FROM N.A.
 RC TISSUE-KIDNEY CORTEX.
 RX MEDLINE: 98191309.
 RA KAUSHAL G.P., SINGH A.B., SHAH S.V.;
 RT "Identification of gene family of caspases in rat kidney and altered
 RT expression in ischemia-reperfusion injury.";
 RL Am. J. Physiol. 274:F587-F595(1998).
 RN [2]
 RP SEQUENCE OF 131-312 FROM N.A.
 RC TISSUE-Ovary;
 RX MEDLINE: 96042508.
 RA FLAWS J.A., KUDOV K., TREBOVICH A.M., DESANTO A., TILLY K.I.,
 RA HIRSHFIELD A.N., TILLY J.L.;
 RT "Interleukin-1 beta-converting enzyme-related proteases (IRPs) and
 RT mammalian cell death: dissociation of IRP-induced oligonucleosomal
 RT endonuclease activity from morphological apoptosis in granulosa cells
 RT of the ovarian follicle.";
 RL Endocrinology 136:5042-5053(1995).
 CC -1- FUNCTION: INVOLVED IN THE ACTIVATION CASCADE OF CASPASES
 CC RESPONSIBLE FOR APOPTOSIS EXECUTION. MIGHT FUNCTION BY EITHER
 CC ACTIVATING SOME PROTEINS REQUIRED FOR CELL DEATH OR INACTIVATING
 CC PROTEINS NECESSARY FOR CELL SURVIVAL (BY SIMILARITY).
 CC -1- SUBUNIT: HETERODIMER OF A SMALL AND A LARGE SUBUNIT (BY
 CC SIMILARITY).
 CC -1- PTM: THE MATURE PROTEASE CAN PROCESS ITS OWN PROPEPTIDE, BUT NOT
 CC THAT OF OTHER CASPASES (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14; ALSO KNOWN AS THE
 CC CASPASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF025671; AAB82567.1;
 DR EMBL: U34684; AAC52260.1;
 DR HSSP: P29466; 1IBC.
 DR PROSITE: PS01121; CASPASE_HIS. 1.
 DR PROSITE: PS01122; CASPASE_CYS. 1.
 DR PFAM: PF00619; CARD. 1.
 DR PFAM: PF00655; ICE_P10. 1.
 DR PFAM: PF00656; ICE_P20. 1.
 KW Hydrolase; Thiol protease; Apoptosis; Zymogen.
 FT NON_TER 1
 FT CHAIN 1 266 CASPASE-2 SUBUNIT 1 (BY SIMILARITY).
 FT CHAIN 267 312 CASPASE-2 SUBUNIT 2 (BY SIMILARITY).
 FT ACT_SITE 210 210 BY SIMILARITY.

FT ACT SITE 253 253 BY SIMILARITY.
 RT NON PER 312 312
 SO SEQUENCE 312 AA: 35070 MW: 532B7699 CRC32:
 ICE2.RAT Length: 312 February 14, 2000 08:02 Type: P Check: 8827

1 TLEKRELIQA KGGSESONE LNLNLPKRG QAFDACEAL RETQGHIED
 51 LLTLTSDIQ HILPLSCDY AQVSPSRQMS PALLISSPAY POILMNPYI
 101 MMYVLPEFRW SHALLSTST LQLAYRQSR PRGLALVSN VHTGEKDE
 151 FRSGGDVDT TLVLEKILG YNVAHYLDQT AQEMERLON FAQLPAHRT
 201 DSCIVALLSH GVEGGIYVD GKLLQLOEV RLEPNANCPN LONKRMFEI
 251 QACRDETRR GVDODGKNH AQPGCESD TYKEELMKR LPTSDMICV
 301 YACIKDNAPI RN

11AA_SEQUENCE 1.0 STANDARD: PRT: 207 AA.
 ID ICM3_PSCORE
 AC P10822:
 DT 01-JUN-1989 (Rel. 11, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 15-JUN-1999 (Rel. 38, Last annotation update)
 DE CHYMOTRYPSIN INHIBITOR 3 PRECURSOR (NCI-3)
 OS PSOPHOCARPUS TETRAGONOLOBUS (Gua bean) (Asparagaceae)
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC eumhyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 CC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
 CC Psophocarpus.
 RN [1]
 RN SEQUENCE FROM N.A.
 RA PEYACHOKNAGUL S., MATSUI T., SHIBATA H., HARA S., IKENAKA T.,
 RA OKADA Y., OHNO T.;
 RT "Sequence and expression of the mRNA encoding the chymotrypsin
 RT inhibitor in winged bean (Psophocarpus tetragonolobus (L.) DC.).";
 RL Plant Mol. Biol. 12:51-58(1989).
 RN [2]
 RN SEQUENCE FROM N.A.
 RA HAHU Y., PEYACHOKNAGUL S., UMEMOTO K., SAKATA Y., OHNO T.;
 RA MEDLINE: 92234991.
 RT "Structure and regulated expression of Kunitz chymotrypsin inhibitor
 RT genes in winged bean [Psophocarpus tetragonolobus (L.) DC.]";
 RL J. Biochem. 111:249-258(1992).
 RN [3]
 RN SEQUENCE FROM N.A.
 RA HAHU Y., SAKATA Y., FUKASAWA K., OHNO T.;
 RA MEDLINE: 94122379.
 RT "Ubiquitous nuclear proteins bind to 5' upstream region of major
 RT Kunitz chymotrypsin inhibitor gene in winged bean";
 RL Plant Mol. Biol. 23:1139-1150(1993).
 RN [4]
 RN SEQUENCE OF 25-207.
 RA SHIBATA H., HARA S., IKENAKA T.;
 RA "Amino acid sequence of winged bean (Psophocarpus tetragonolobus (L.)
 RT DC.) chymotrypsin inhibitor, WCI-3.";
 RL J. Biochem. 104:537-543(1988).
 RN [5]
 RN X-RAY CRYSTALLOGRAPHY (2.95 ANGSTROMS).
 RC TISSUE=SEED;
 RA DATAGUPTA J. K., PODDER A., CHAKRABARTI C., SEN U., DUTTA S. K.,
 RA SINGH M.;
 RA "Structure of a Kunitz-Type chymotrypsin from winged bean seeds at
 RT 2.95-A resolution.";
 RL Acta Crystallogr. D 52:521-528(1996).
 RN [6]
 RN X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
 RA MEDLINE: 99258819.
 RA DATAGUPTA J. K., PODDER A., CHAKRABARTI C., SEN U., MUKHOPADHYAY D.,
 RA DUTTA S. K., SINGH M.;

RT RT "Refined crystal structure (2.3 Å) of a double-headed winged bean
 RT alpha-chymotrypsin inhibitor and location of its second reactive
 RT site.";
 RL Proteins 35:321-331(1999).
 RN [7]
 RN X-RAY CRYSTALLOGRAPHY (2.13 ANGSTROMS).
 RA RAVICHANDRAN S., SEN U., CHAKRABARTI C., DATAGUPTA J. K., DUTTA S. K.,
 RA SINGH M.;
 RA "2.13-Å structure of a Kunitz-type winged bean chymotrypsin inhibitor
 RT protein." (Oct-1998) to the PDB data bank.
 RT Submitted: INHIBITS ALPHA-CHYMOTRYPSIN AT THE MOLAR RATIO OF 1:2
 CC -1- FUNCTION: INHIBITS ALPHA-CHYMOTRYPSIN AT THE MOLAR RATIO OF 1:2
 CC IN STATE OF 1:1.
 CC -1- SIMILARITY: TO OTHER LEGUMINOUS KUNITZ-TYPE INHIBITORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: S96732; AAC60535.1; -
 DR EMBL: D13974; BAA03084.1; -
 DR EMBL: D13975; BAA03085.1; -
 DR EMBL: D13976; BAA03086.1; -
 DR PIR: JX0206; JX0206.
 DR PIR: JX0206; JX0206.
 DR PIR: S42563; S42563.
 DR PIR: S42564; S42564.
 DR PDB: 1WBC; 03-APR-96
 DR PDB: 2WBC; 25-FEB-98
 DR PDB: 4WBC; 12-MAR-99.
 DR PROSITE: PS00283; SOYBEAN_KUNITZ; 1.
 DR PFM: PF00197; Kunitz_legume; 1.
 DR Serine protease inhibitor; Signal: 3D-structure.
 FT SGNAL 1 24
 FT CHAIN 25 207 CHYMOTRYPSIN INHIBITOR 3.
 FT DISULFID 65 109
 FT DISULFID 159 168
 FT ACT_SITE 89 90
 SO SEQUENCE 207 AA: 22774 MW: 2288B770 CRC32:
 ICM3_PSCORE Length: 207 February 14, 2000 08:02 Type: P Check: 7697

1 MKSTFLALF LLSALISHLP SSTADDLVD AEGNLVENG TYLLPRHWA
 51 HGGIEETAKT GNEPCPLTV RSPNEVSKGE PRISSEFLS LFIROSLVA
 101 LGFANPPSCA ASPWTIVDS PQGPAVKLSQ QKLPKEDLV EFKFKVSHSN
 151 INHYKLLYCO HHEEDVKCDQ YIGIHDRNG NRRLVVTEN PLEIVILKAK
 201 SETASSH

11AA_SEQUENCE 1.0 STANDARD: PRT: 203 AA.
 ID ID1ACTV
 AC P7202; 008150.
 DT 15-DEC-1999 (Rel. 39, Created)
 DT 15-DEC-1999 (Rel. 39, Last sequence update)
 DT 15-DEC-1999 (Rel. 39, Last annotation update)
 DE PUTATIVE ISOPENTENYL-DIPHOSPHATE DELTA-ISOMERASE (EC 5.3.3.2) (IPE
 DE ISOMERASE) (ISOPENTENYL PYROPHOSPHATE ISOMERASE).
 GN RV1745C OR MTCY28.08C OR MTCY04C12.29C.
 OS Mycobacterium tuberculosis.
 CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 CC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 RN [1]
 RN SEQUENCE FROM N.A.
 RA STRAIN-H37RV;
 RX MEDLINE: 98295987.
 RA COLE S. T., BROCH R. J., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
 RA GORDON S. V., EIGLMEIER K., GAS S., BARRY C. E. III, TEKAIA F.,

RA BADOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
 RA DAVIES R., DEVLIN K., FELTWEILL T., GENTLES S., HAMLIN N., HOLROYD S.,
 RA HORNBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
 RA OLIVER S., OSBORNE J., OUALI M.A., RAJANDREEM M.A., ROGERS J.,
 RA RUTTER S., SEEGER K., SKELTON S., SQUARES S., SQUARES R., STILSON J.E.,
 RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 CC -1- FUNCTION: CATALYZES THE 1,3-ALLYLIC REARRANGEMENT OF THE
 CC HOMALYLIC SUBSTRATE ISOPENTENYL (IPP) TO ITS ALLYLIC ISOMER,
 CC DIMETHYLLYL DIPHOSPHATE (DMAPP) (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ISOPENTENYL DIPHOSPHATE = DIMETHYLLYL
 CC DIPHOSPHATE.
 CC -1- SIMILARITY: BELONGS TO THE IPP ISOMERASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: 295890; CAB09331.1; -
 KW Hypothetical protein; Isomerase; Isoprene biosynthesis.
 FT ACT_SITE 75 75 BY SIMILARITY.
 FT ACT_SITE 124 124 BY SIMILARITY.
 SO SEQUENCE 203 AA; 22489 MW; 4D4FFD37 CRC32;
 IDL_MYCTU Length: 203 February 14, 2000 08:02 Type: P Check: 9024 ..
 1 MRSYRPAAP IERYVLNDR GDAIVADKA TVHTGDFPLH LARSSYFDE
 51 HQGLITRRA AFRTWPAVM TNSCGHPLE GESLPQAIR RLAAELGLTP
 101 DRYDLIPGF RYRAMADGT VENEICPYR VQVDQPRPN SDEVDAIRML
 151 SMEQFVRDVT AGVAPVSPW CRSQGLYLR LGPCRAQMPV ADCCRPLKKA
 201 HGN
 !!AA SEQUENCE 1.0
 ID IDNR_ECOLI STANDARD: PRT; 332 AA.
 AC P09343;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-DEC-1999 (Rel. 39, Last annotation update)
 DE L-IDONATE REGULATORY PROTEIN.
 GN IDNR.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE: 953434362.
 RA BURLAND V.D., PLUNKETT G. III, SOFIA H.J., DANIELS D.L.,
 RA BLATTNER F.R.;
 RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
 RT region from 92.8 through 100 minutes."
 RL Nucleic Acids Res. 23:2105-2119(1995).
 RN [2]
 RP FUNCTION: 98324983.
 RA BAUSCH C., PEKHAUS N., UTZ C., BLAIS T., MURRAY E., LOWARY T.,
 RA CONWAY T.;
 RT "Sequence analysis of the gntII (subsidiary) system for gluconate
 RT metabolism reveals a novel pathway for L-idonic acid catabolism in
 RT Escherichia coli."
 RL J. Bacteriol. 180:3704-3710(1998).
 CC -1- FUNCTION: IDN OPERON REGULATOR. MAY REPRESS GNTKU AND GNTT GENES
 CC WHEN GROWING ON L-IDONATE.

CC -1- SIMILARITY: BELONGS TO THE LACI FAMILY OF TRANSCRIPTIONAL
 CC REGULATORS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: U14003; AA097161.1; -
 CC EMBL: AE000497; AC77221.1; -
 DR HSSP; P03023.1ICD; IDNR.
 DR ECGENE; EG12538; IDNR.
 DR PROSITE; PS00356; HTH_LACI_FAMILY; 1.
 DR PFAM; PF00356; Laci; 1.
 KW Transcription regulation; DNA-binding; Gluconate utilization.
 FT DNA_BIND 8 27 H-T-H MOTIF (POTENTIAL).
 SO SEQUENCE 332 AA; 37567 MW; B4E1F229 CRC32;
 IDNR_ECOLI Length: 332 February 14, 2000 08:02 Type: P Check: 2843 ..
 1 MRNRISLQD IATLAVTKM TVSRYIRSP KYAKETGERI AKIMEINYI
 51 PNRAFGMLN AGSYTLGILI PSFQNLFAD ILAGIESVTS EHYQTILAN
 101 YNYDRDSEEE SYVNLSTYNI DGIILSEKH TIRTVFLNS ATIVVELMD
 151 VQGRIDMEV GFQNRQAARD MVTMLKRY RHIIYLGSK DQTRDEQRYO
 201 GYCAMMLHN LSLRLNPPA ISSIHLNQL MDALSNAND LDGFCINDD
 251 IANALLLCR ERLAVPEOI SIAGFGLER GRQMPSLAS VITRPDGR
 301 MAQMLSKI KNDNHNTV DLGYQYHGN TL
 !!AA SEQUENCE 1.0
 ID IDNR_ECOLI STANDARD: PRT; 131 AA.
 AC P09182;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE COLICIN N IMMUNITY PROTEIN (MICROCIN N IMMUNITY PROTEIN).
 GN CNI.
 OS Escherichia coli.
 OC Plasmid Coln pCHAP4.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE: 88174431.
 RA PUGSLEY A.P.;
 RT "The immunity and lysis genes of Coln plasmid pCHAP4."
 RL Mol. Gen. Genet. 211:335-341(1988).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: X06933; CA030020.1; -
 DR PIR; S01761; IMECN4.
 KW Plasmid; Bacteriocin; Colicin; Transmembrane.
 FT TRANSMEM 66 84 POTENTIAL.
 FT TRANSMEM 104 124 POTENTIAL.
 SO SEQUENCE 131 AA; 15245 MW; BA839D94 CRC32;

IMM_ECOLI Length: 131 February 14, 2000 08:02 Type: P Check: 6528 ..

1 MHNTLEKIT ATLSLPGFHS LNNPPLSEAF NLXYHTAPLA ATSLFITHK

51 ELEIKPRSSP LPAKILITPE TILYISMITC FLITDTLELT SSKTFVLIVK

101 KRSEVEFFLY NTIWDIYH IFVLIVRYN I

11AA_SEQUENCE 1.0 STANDARD; PRT; 424 AA.

ID IMPB_SALTY

AC P18642;

DT 01-NOV-1990 (Rel. 16, Created)

DT 01-NOV-1990 (Rel. 16, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE IMPB PROTEIN.

GN IMPB.

OS Salmonella typhimurium.

OG Plasmid Incil TP110.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

CC Salmonella.

CC [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 9038479.

RA LODWICK D., OWEN D., STRIKE P.

RT "DNA sequence analysis of the imp UV protection and mutation operon

of the plasmid p110: identification of a third gene."

Nucleic Acids Res. 18:5045-5050(1990).

CC -1- FUNCTION: INVOLVED IN UV PROTECTION AND MUTATION.

CC -1- SIMILARITY: BELONGS TO THE IMPB/MOCE/SABE FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

CC -----

DR EMBL: X53528; CAA37608.1; -

DR PIR: J00661; J00661.

DR PIR: J00661; J00661.

DR PIR: J00661; J00661.

DR PIR: J00661; J00661.

DR PIR: J00661; J00661.

DR PIR: J00661; J00661.

DR PIR: J00661; J00661.

DR PIR: J00661; J00661.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida;

OC Alismataceae; Sagittaria.

CC [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 93224493.

RA XU W., TAO W., GONG Z., CHI C.-W.;

RT "CDNA and genomic structures of arrowhead proteinase inhibitors."

J. Biochem. 113:153-158(1993).

CC [2]

RP SEQUENCE OF 25-174.

RX MEDLINE: 92316904.

RA YANG H.-L., LUO R.-S., WANG L.-X., ZHU D.-X., CHI C.-W.;

RT "Primary structure and disulfide bridge location of arrowhead double-

headed proteinase inhibitors."

J. Biochem. 111:537-545(1992).

CC -1- FUNCTION: POSSESSES TWO REACTIVE SITES. INHIBITS AN EQUIMOLAR

AMOUNT OF TRYPSIN AND CHYMOTRYPSIN SIMULTANEOUSLY, AND INHIBITS

KALLIKREIN WEAKLY.

CC -1- SIMILARITY: STRONG, TO S. SAGITTIFOLIA PROTEINASE INHIBITORS B.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

CC -----

DR EMBL: D13819; BA02972.1; -

DR PIR: A13809; AKRA.

DR PIR: J00246; J00246.

DR PIR: J00246; J00246.

DR PIR: J00246; J00246.

DR PIR: J00246; J00246.

DR PIR: J00246; J00246.

DR PIR: J00246; J00246.

DR PIR: J00246; J00246.

DR PIR: J00246; J00246.

DR PIR: J00246; J00246.

DR PIR: J00246; J00246.

DR PIR: J00246; J00246.

DR PIR: J00246; J00246.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida;

OC Alismataceae; Sagittaria.

CC [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 93224493.

RA XU W., TAO W., GONG Z., CHI C.-W.;

RT "CDNA and genomic structures of arrowhead proteinase inhibitors."

J. Biochem. 113:153-158(1993).

CC [2]

RP SEQUENCE OF 25-174.

RX MEDLINE: 92316904.

RA YANG H.-L., LUO R.-S., WANG L.-X., ZHU D.-X., CHI C.-W.;

RT "Primary structure and disulfide bridge location of arrowhead double-

headed proteinase inhibitors."

J. Biochem. 111:537-545(1992).

CC -1- FUNCTION: POSSESSES TWO REACTIVE SITES. INHIBITS AN EQUIMOLAR

AMOUNT OF TRYPSIN AND CHYMOTRYPSIN SIMULTANEOUSLY, AND INHIBITS

KALLIKREIN WEAKLY.

CC -1- SIMILARITY: STRONG, TO S. SAGITTIFOLIA PROTEINASE INHIBITORS B.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

CC -----

DR EMBL: D13819; BA02972.1; -

DR PIR: A13809; AKRA.

DR PIR: J00246; J00246.

DR PIR: J00246; J00246.

DR PIR: J00246; J00246.

DR PIR: J00246; J00246.

DR PIR: J00246; J00246.

DR PIR: J00246; J00246.

DR PIR: J00246; J00246.

DR PIR: J00246; J00246.

DR PIR: J00246; J00246.

DR PIR: J00246; J00246.

DR PIR: J00246; J00246.

DR PIR: J00246; J00246.

RT system of Streptomyces albus G.;

RL Mol. Genet. 225:142-147(1991).

CC -1- FUNCTION: INACTIVATION OF SAL I RESTRICTION-MODIFICATION SYSTEM.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: X56644; CAA39965.1; -

KM Transposable element.

SO SEQUENCE 256 AA; 28409 MW; AD5CD17C CRC32;

IS12_STRAL Length: 256 February 14, 2000 08:02 Type: P Check: 4079 ..

1 MAGVITASEP SWIAPFSGLS PROFGKLVTV LRREGADAVR KGRFWSIPLE

51 DRALLVAAYV RNTLTMROLA PLEGVSKSAA DRIIDHIGPM LALQPRKRF

101 KDVLIVDGT LVPTRDHTIA ERSKNRYST NQAVIADDT RLIVVYGRPL

151 AGNRNCKAV EESGAKAAGV KTLIADGKY PGTGLVTPHR RENQAGLPD

201 WKEHNKSHK QVAREVEHF ARMTKILR DCLRGDGVH HAMLGIRMH

251 NLALTG

II.AA SEQUENCE 1.0

ID ISIA.SYNY3 STANDARD: PRT: 342 AA.

AC 055274; P73884;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE IRON-STRESS INDUCED CHLOROPHYLL-BINDING PROTEIN.

GN ISIA OR SL0247.

OS Synchocystis sp. (strain PCC 6803).

OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.

RN (1)

RA SEQUENCE FROM N.A.

RA FERRIERA F., STRAUS N.A.;

RL Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.

RN (2)

RA SEQUENCE FROM N.A.

RX MEDLINE; 97061201.

RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,

RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAKOTO S., KIMURA T.,

RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NABURO K.,

RA OKUMURA S., SHIMO S., TAKEUCHI C., WADA T., WATANABE A.,

RA YAMADA M., YASUDA M., TABATA S.;

RT "Sequence analysis of the genome of the unicellular cyanobacterium

RT Synchocystis sp. strain PCC6803. II. Sequence determination of the

RT DNA Res 3:109-136(1996).

RL -1- FUNCTION: PHOTOSYSTEM II CHLOROPHYLL-BINDING PROTEIN.

CC -1- INDUCTION: BY IRON STRESS.

CC -1- SIMILARITY: BELONGS TO THE PSBB / PSBC FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: L26530; AAA27291.1; -

DR EMBL: D90910; BAA17948.1; -

DR PFAM: PF00421; PSII.2;

KW Photosystem II; Chlorophyll; Transmembrane.

FT TRANSMEM 30 51 I.

FT DOMAIN 63 81 II. VERY HYDROPHOBIC.

FT TRANSMEM 89 113 III.

FT TRANSMEM 141 167 IV.

FT TRANSMEM 204 223 V.

FT TRANSMEM 240 262 VI.

FT TRANSMEM 322 323 VII. VERY HYDROPHOBIC.

FT CONFLICT 94 94 S -> R (IN REF. 1).

FT CONFLICT 334 342 EDAFDSLOT -> NKRLIPCKPSH (IN REF. 1).

SO SEQUENCE 342 AA; 37221 MW; 889359C7 CRC32;

ISIA.SYNY3 Length: 342 February 14, 2000 08:02 Type: P Check: 4153 ..

1 MOTYGNDTVQ YEWMAGNARF ADQSGFLTA HVAQAALTAE WAGATLFEI

51 SNFDTQAMG DOGLILPHL ATLGVGVGDG GQIVDTYPYF VIGSIHLIAS

101 AVLGAGALFH TLRAPADLST LKGGKKRHF TWENPOOLGI ILGHLLFLG

151 AGALLAGKA MYWGLYDAT TOTVATVQSP TLDPVTVGY QTFASISSL

201 EDLVGHHFV GFLIGGIM HILVPLGMA KVLLESGBA ILSYSLGIA

251 LAGFAAVFC AVNTLAPPE FYGPPIAKL GIPPEADIV ELPMHAHSR

301 AMLANHFEL APFLDGHLM HALRALGTFD KRYEQAFLSL QT

II.AA SEQUENCE 1.0

ID KDUD.BACST STANDARD: PRT: 254 AA.

AC P50842;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE 2-DEOXY-D-GLUCONATE 3-DEHYDROGENASE (EC 1.1.1.125) (2-KETO-3-DEOXYGLUCONATE OXIDOREDUCTASE).

GN KDUD.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/staphylococcus group; Bacillus.

RN (1)

RA SEQUENCE FROM N.A.

RA STRAIN-168 / MARBURG;

RX MEDLINE; 96349105.

RA SOROKIN A.V., AZEVEDO V., ZUMSTEIN E., GALLERON N., EHRLICH S.D.,

RA SERROR P.;

RT "Sequence analysis of the Bacillus subtilis chromosome region between

RT the sera and kgd loci cloned in a yeast artificial chromosome.";

RL Microbiology 142:2005-2016(1996).

CC -1- CATALYTIC ACTIVITY: 2-DEOXY-D-GLUCONATE + NAD(+) = 3-DEHYDRO-2-DEOXY-D-GLUCONATE + NADH.

CC -1- PATHWAY: PEPTIN DEGRADATION

CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES FAMILY (SDR).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: L47838; AAB38476.1; -

DR EMBL: Z99115; CAB1431.1; -

DR HSP: P19992; ZHSD.

DR SUBTILIST: BG11400; KDUD.

DR PROSITE: PS00061; ADH.SHORT.1.

DR PFAM: PF00106; adh.short.1.

KW Oxidoreductase; NAD.

FT NP_BIND 16 40 NAD (BY SIMILARITY).

FT ACT SITE 159 159 BY SIMILARITY

SO SEQUENCE 254 AA; 27185 MW; 44F5CB25 CRC32;

KDD_BACSU Length: 254 February 14, 2000 08:02 Type: P Check: 3484

1 MG1HDAFSL KGTALVTGP GTGIGIGIAK ALAGAGADII GTSHTSLSE
51 TOOLVQEGSR IFTSFILDMS KPEAIKDSAA ELFENROIDI LVNNAGIIHR
101 EKAEDPEEN WQVHLNVNLSL IFTILQOLAG RHMILKRGHK IINIASLST
151 OGGLIVPATY ASKHAVAGLT KSPANWMAAS GIOVVALIAG YISANTKPI
201 RODEKRNEDI LKRIPAGRW QADIDIGTAV FLASASDXY NGHILAVDGS
251 WLSR

11AA_SEQUENCE 1.0 STANDARD: PRT: 467 AA.
ID KRL_HSVBS
AC 008097;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE SERINE/THREONINE-PROTEIN KINASE (EC 2.7.1.-).
OS Bovine herpesvirus type 1.2 (strain ST).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicelloviruses.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94167875.
RA LEUNG-TRICK P., AUDONNET J.F., RIVIERE M.;
RT "The complete DNA sequence and the genetic organization of the short
RT unique region (US) of the bovine herpesvirus type 1 (ST strain)."
RL Virology 199:409-421(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: 223068; CA80602.1;
DR HSSP: P24941; IAOI.
DR PROSITE: P500107; PROTEIN_KINASE_ATP_FALSE_NEG.
DR PROSITE: P500108; PROTEIN_KINASE_ST; 1.
DR PROSITE: P50011; PROTEIN_KINASE_DOM; 1.
DR PFAM: PF00069; PKinase; 1.
KW transferase; Serine/threonine-protein kinase; ATP-binding.
FT DOMAIN 164 462
FT NP_BIND 170 178
FT BINDING 194 194
FT ACT_SITE 279 279
FT ACT_SITE BY SIMILARITY.
SQ SEQUENCE 467 AA: 49885 MW: 255B1E22 CRC32;

KRL_HSVBS Length: 467 February 14, 2000 08:02 Type: P Check: 5295

1 MERAARLRL QRAIGLMSR FACCVAAEPS GSRLGSGVNG AAAAPARCAA
51 EGSADILYLV NNEGPEVAP ARTGPPDAG IEGGAAYVGN EGGVAAGNE
101 RRAATGDENE SASGGENESE SESSESESE SGADGDMD DDDAGPAGV
151 TREBAGGAR ALNRIIRRL TPSEGEVFE ATGPAPQDH VVAKIGASAS
201 TLAEMILRT LDHANVYKL AVLEFGEIYV AVLARYREDL HTLMKIDRP
251 MALPTALQVT RAVILOGLAYL HSRRIARHV KTEENFLNGP GDVCLGDFGA
301 AHGPVTEPRY YGLAGTLETN SPELLARARY DCRIDVMSAG VVAEMLAYP
351 RALEDSPAGP OGEDAASGP PTILGRDCA ROLLRYIRRL AVHAEEPPPS
401 PTDRLTRNFK RHAATGREPH SPYRCIAVLR LPCDADRLHL QMLTFERRAR

451 PTAELELHP VEGASAG

11AA_SEQUENCE 1.0 STANDARD: PRT: 202 AA.
ID LSPA_MYCTU
AC 010764;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE PROBABLE LIPOPROTEIN SIGNAL PEPTIDASE (EC 3.4.23.36) (PROLIPOPROTEIN
DE SIGNAL PEPTIDASE) (SIGNAL PEPTIDASE II) (SPASE II).
GN LSPA OR RV1539 OR MYCY48.26C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriinae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 98265987.
RX STRAIN-H37RV;
RA COLE S.T., BROSCH R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA GORDON S.V., EIGMEYER K., GINS S., BARRY C.E. III, TERKLA E.,
RA BADCOCK K., BASHAM D., BROWN D., CHILINCWORTH H., CONNOR R.,
RA DAVIES R., DEVLIN K., FELTWEIL T., GENTLES S., HAMLIN N., HOLROYD S.,
RA HORSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
RA OLIVER S., OSBORNE J., QUAIL M.A., RAUNDREEM M.A., ROGERS J.,
RA RUTTER S., SEGER R., SKELTON S., SQUARES S., SQUARES R., SULLSTON J.E.,
RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
CC -1- FUNCTION: THIS PROTEIN SPECIFICALLY CATALYZES THE REMOVAL OF
CC SIGNAL PEPTIDES FROM PROLIPOPROTEINS.
CC -1- CATALYTIC ACTIVITY: CLEAVAGE OF N-TERMINAL LEADER SEQUENCES FROM
CC MEMBRANE PROLIPOPROTEINS. HYDROLYSES XAA-XBB-XBB-1-CYS, IN WHICH
CC XAA IS HYDROPHOBIC (PREFERABLY LEU), XBB IS OFTEN SER OR ALA, XCC
CC IS OFTEN GLY OR ALA, AND THE CYS IS ALKYLATED ON SULFUR WITH A
CC DIACYLGYCERYL GROUP.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: 274020; CA98323.1;
DR PROSITE: P500855; SPASE_II; 1.
DR PFAM: PF01252; Peptidase_A8; 1.
KW Hypothetical protein; Hydrolase; Aspartyl protease; Transmembrane.
FT DOMAIN 1 34
FT TRANSMEM 35 55
FT DOMAIN 56 87
FT TRANSMEM 88 108
FT DOMAIN 109 111
FT TRANSMEM 112 132
FT DOMAIN 133 159
FT TRANSMEM 160 180
FT DOMAIN 181 202
FT TRANSMEM 203 213
FT TRANSMEM 214 234
FT TRANSMEM 235 254
SQ SEQUENCE 202 AA: 21345 MW: 906DB13F CRC32;

LSPA_MYCTU Length: 202 February 14, 2000 08:02 Type: P Check: 2319

1 MPDEPGSND PLTTEBAGG AGEPNAPAP RRLMLLSVA VVVLTIDIVT
51 KYVAQVLDP GPVSIIGDT VVITLVVNSG AAFSNATGYT WVLTLATGV
101 VVGIFWGR LVSPWMAIGL GMTLGGAMN LVDRFFRAGP PLRGHVVDFL
151 SVGWPEVENV ADPSVVGAI LVIYLSIFG DDTVGRRA DGTVGRRA
201 DG

11AA_SEQUENCE 1.0 STANDARD; PRT; 277 AA.
 ID MAUF_PARDE
 AC P29897;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE METHYLAMINE UTILIZATION PROTEIN MAUF.
 GN MAUF.
 OS Paracoccus denitrificans.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
 OC Paracoccus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DSM 413;
 RX MEDLINE; 95045590.
 RA VAN SPANING R.J.M., DER PALEN C.J., SLOTBOOM D., REIJNDERS W.N.,
 RA STOUTRAMER A.H., DUINE J.A.;
 RT "Expression of the mau genes involved in methylamine metabolism in
 RT Paracoccus denitrificans is under control of a Lysr-type
 RT transcriptional activator."
 RL Eur. J. Biochem. 226:201-210(1994).
 RN [2]
 RP SEQUENCE OF 129-277 FROM N.A.
 RX MEDLINE; 92272712.
 RA CHRISTENSEN D.A.Y., BOYD J., MATHENS F.S., LIDSTROM M.E.;
 RT "A transcription-dependent DnaE I hypersensitive site in a far
 RT upstream segment of the chicken alpha-globin gene domain coincides
 RT with a matrix attachment region."
 RL Biochem. Biophys. Res. Commun. 184:1181-1189(1992).
 CC -1- PATHWAY: METHYLAMINE UTILIZATION.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U12464; AAA56723.1;
 DR EMBL; M90099; AAA25579.1;
 DR PIR; PH0856; PH0856.
 KW Transmembrane.
 FT TRANSMEM 33 53 POTENTIAL.
 FT TRANSMEM 59 79 POTENTIAL.
 FT TRANSMEM 111 131 POTENTIAL.
 FT TRANSMEM 132 152 POTENTIAL.
 FT TRANSMEM 176 196 POTENTIAL.
 FT TRANSMEM 205 225 POTENTIAL.
 FT TRANSMEM 257 277 POTENTIAL.
 SQ SEQUENCE 277 AA; 28361 MW; 7BE517A2 CRC32;
 MAUF_PARDE Length: 277 February 14, 2000 08:02 Type: P Check: 7474 ..

DE MALATE DEHYDROGENASE (EC 1.1.1.37).
 GN MDH
 OS Rhizobium leguminosarum (biovar viciae)
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-3841;
 RA POOLE P.S., ALLANBY D., SMITH M.;
 RL Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: L-MALATE + NAD(+) = OXALOACETATE + NADH.
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AJ002750; CA05717.1;
 DR HSSP; Q27743; 1LDG.
 DR PROSITE; PS00068; MDH; FALSE_NEG.
 DR PIR; PF00056; 1db; 1.
 KW Oxidoreductase; Tricarboxylic acid cycle; NAD.
 FT ACT_SITE 149 149 PROTON-RELAY (BY SIMILARITY).
 FT BINDING 152 152 SUBSTRATE CARBOXYL GROUP (BY SIMILARITY).
 FT ACT_SITE 176 176 PROTON-RELAY (BY SIMILARITY).
 SQ SEQUENCE 320 AA; 35590 MW; 192E3E24 CRC32;
 MDH_RHIV Length: 320 February 14, 2000 08:02 Type: P Check: 1419 ..

1 MAANKIALIG SCMIGTIAH LAQIKELGDI VLFDIADGIP QGKGLDISQS
 51 SPVEGDEVNL TGASDYSAIE GADVCIYTAG VARKGMSRD DLIGINLKV
 101 EGVAGIKIKY APNAFVICIT NPIDAWMAL QKFSGLPANK VGVNAGVLDS
 151 SRRFLAKE FNVSYODVTA FVLGHDGM VPLARYSTVG GIPLTDLVTM
 201 GWTKERLEE IIOFRDGA EIVGLKTGS AYAFAASAI EMASYLKDK
 251 KVLPCAAH SGQGVKDMY VGVPIYVIGAG GVERIIEIDL NTKKEAFDK
 301 SVGAVAGLCE ACINIPALK

11AA_SEQUENCE 1.0 STANDARD; PRT; 430 AA.
 ID MDH_RHIV
 AC P4689;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE PUTATIVE CELL CYCLE PROTEIN MESJ HOMOLOG.
 GN MESJ OR H10404.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RD / KW20;
 RX MEDLINE; 95350630.
 RA FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
 RA KERLAVAGE A.R., BULL C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
 RA MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
 RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODER A., KELLEY J.M.,
 RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,
 RA UTTERBACK T.R., HANNA M.C., NGUYEN P.T., SAUDEK D.M., BRANNON R.C.,
 RA FINE L.D., FRITZMAN J.L., FUHRMANN J.L., GEOGAGEN N.S.M.,
 RA GNEH C.L., MCDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
 RA VENTER J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus

RT Influenzae Rd. " Science 269:496-512(1995).
 CC -1- SIMILARITY: BELONGS TO THE UPF0072 (MESJ/YCF62) FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U32723; AAC22063.1; -
 DR TIGR: W10404;
 DR SEQUENCE 430 AA; 50052 MW; BE9205D2 CRC32;
 SQ
 MESJ_HAEIN Length: 430 February 14, 2000 08:02 Type: P Check: 1410 ..
 1 MDLSDIERO LOKATAQAFI IASGGIDST VLSLFKIC OKOPHLPLS
 51 IRAIHINGL SPNADSWAKH CODICDOFOI PLIERVOVD KTNCEIAGAR
 101 EARYOAIKKY IOTOEMLYTA HMLNDQETI FLAKRSGI KGLGMOQOS
 151 VLFQMLIRP ILGFTTOLY NYAQEKILW ITDESNDNR YDRNFLNEI
 201 LPELRERMAH FDLAVORSNO HCFEODOLIN DLLSEIFTEH COIKNOERLC
 251 QFRQYSIAKQ TALLRWMLAE NOLEMPKROQ LTOLINDYIF AKERANPOFO
 301 LVNKVIRRYO DSYLTKEPFS DLTKCTKLE QNTLNPDOL GNLFVQENEH
 351 NLFFWQDYS VLEKTNLPI SIRGYSKGV KHYPRPRPD IKKIWOELGV
 401 PPMERNRIPL IFYGNELSKA VGFEFVRLKSS
 11AA SEQUENCE 1.0
 ID MPXA MYXKA STANDARD; PRT: 236 AA.
 AC P38371;
 DT 01-OCT-1994 (rel. 30, Created)
 DT 01-OCT-1994 (rel. 30, Last sequence update)
 DT 01-FEB-1995 (rel. 31, Last annotation update)
 DE PUTATIVE LIPOPROTEIN MLPA PRECURSOR.
 GN MLPA.
 OS Myxococcus xanthus.
 OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
 CC Myxococcales; Cystobacteriineae; Myxococcaceae; Myxococcus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DZFI;
 RX MEDLINE: 93328680.
 RA MARTINEZ-CANAMERO M., MUNOZ-DORADO J., FAREZ-VIDAL E., INOUE M.,
 RA INOUE S.;
 RT "Car. a 115-kilodalton membrane protein required for development of
 RT Myxococcus xanthus.";
 RT J. Bacteriol. 175:4756-4763(1993).
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID
 CC ANCHOR (PROBABLE).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: S64103; AAB27615.1; -
 DR PIR: B40609; B40609.
 DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 KW Membrane; Signal; Lipoprotein.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 236 PUTATIVE LIPOPROTEIN MLPA.

FT LIPID 22 22 N-ACYL DIGLYCERIDE (POTENTIAL).
 SQ SEQUENCE 236 AA; 24618 MW; 97CAB418 CRC32;
 MPXA MYXKA Length: 236 February 14, 2000 08:02 Type: P Check: 8809 ..
 1 MKRNIVNTAL VLVAGSILIT GCNFEOPETN CFVQSPSWA VKYVDVDSRK
 51 DANGDECTIT APVLKMGV KYVNPETGAA QIALRPATIA SRAIDTITTT
 101 SADQISIGSL DTEPKDHGFC HANDEPAFV NVAASDTRAA NTIRYEFTNV
 151 RYSAVAAPG TOFTGELKYT SNGCTSSYVA RAWPPAPCD TASTEPAENC
 201 GVGSGLNPEF AVVQPSAT GTTGCVCPAG DIPSRK
 11AA SEQUENCE 1.0
 ID MML6 MYCTU STANDARD; PRT: 397 AA.
 AC Q10773;
 DT 01-OCT-1996 (rel. 34, Created)
 DT 01-OCT-1996 (rel. 34, Last sequence update)
 DT 15-DEC-1999 (rel. 39, Last annotation update)
 DE PUTATIVE MEMBRANE PROTEIN MML6.
 GN MML6 OR RV1357 OR MTC148.08C.
 OS Mycobacterium tuberculosis.
 CC Bacteria; Firmicutes; Actinobacteria; Actinobacteriidae;
 CC Actinomycetales; Corynebacteriineae; Mycobacteriaceae; Mycobacterium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RX MEDLINE: 98295987.
 RA COLE S.T., BROSC R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
 RA GORDON S.V., EIGMEIER K., GAS S., BARRY C.E. II, TEKAIA F.,
 RA BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
 RA DAVIES R., DEVLIN K., FETTWELL T., GENTLES S., HAMLIN N., HOLROYD S.,
 RA HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
 RA OLIVER S., OSBORNE K., QUAIL M.A., RAJANDREAN M.A., ROGERS J.,
 RA RUTTER S., SEGER K., SKELTON S., SQUARES S., SULLSTON J.E.,
 RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RT Nature 393:537-544(1998).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z74020; CAA98334.1; -
 KM Hypothetical protein; Transmembrane.
 FT TRANSMEM 161 181 POTENTIAL.
 FT TRANSMEM 210 210 POTENTIAL.
 FT TRANSMEM 214 234 POTENTIAL.
 FT TRANSMEM 242 262 POTENTIAL.
 FT TRANSMEM 293 313 POTENTIAL.
 FT TRANSMEM 330 350 POTENTIAL.
 SQ SEQUENCE 397 AA; 42421 MW; 9ARDFAD0D CRC32;
 MML6 MYCTU Length: 397 February 14, 2000 08:02 Type: P Check: 4418 ..
 1 MGIISTGVY KRGWVRYSF DTIDIGIQIG EQLASTYVIL DKLAIQPOL
 51 VALPDEIAS QOINRELAIA NYATMSGIYA QNALIENNA AMGAPDAK
 101 NDSFYLPE AFDPDFORG LKFLSADGK AARMTISHEG DPATPEGISH
 151 IDAIKAANE AVKGTPMAGA GIYIAGTAAT FPDIDGATY DLLIAGTAAL
 201 SLILIMMI TRSLVAALVI VGTVALSIGA SEGLSVLWQ HLLGIOLYWI

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
 CC EMBL: X12599; CAA31114.1; -
 CC EMBL: X13303; CAA31670.1; -
 CC PIR: S01838; S01838.
 DR Nitrogen fixation.
 KW CONFLICT 170 170 L -> P (IN REF. 2).
 FT CONFLICT 208 229 GLTGITLKNAFENPCFVLVAR -> RLNRNDPLEERYV
 FT (IN REF. 2).
 SQ SEQUENCE 229 AA; 25517 MW; 4F5A88DE CRC32;
 NITR_KLEPN Length: 229 February 14, 2000 08:02 Type: P Check: 8799 ..

1 MSNDNLEFMR MIALFQSLPD LQPAQIVDWL AGESETLTP ERLATLTPQ
 51 LAASFSRATA VMSPARMSRV MASLQGLPA HLRITRPNOR TPOLLAFC
 101 QDGLVINGHF GGGRLFTIYA FDEGGWLYD LRRYSAPHQ QEAENEVARR.
 151 IEDCQLLEFC EIGGPAARL IRRIRHPKA QPCTTQAC EAINVLLAGR
 201 LPPWLAKGLT GITLWKNAF NPCFVLVAR

11AA_SEQUENCE 1.0
 ID NORL_ASPPA STANDARD; PRT; 271 AA.
 AC Q00278;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE AFLATOXIN BIOSYNTHESIS KETOREDUCTASE NOR-1 (EC 1.1.1.-).
 OS NOR-1 OR NAR-1.
 GN Aspergillus parasiticus.
 CC Eukaryote; Fungi; Ascomycota; Euascomycetes; Plectomycetes;
 CC Eutotals; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 CC [1]
 CC SEQUENCE FROM N.A.
 RP STRAIN:ATCC 163 / NRRL 5862 / SU-1;
 RC MEDLINE; 95085270
 RA TRAIL F., CHANG P.-K., GARY J., LINZ J.E.;
 RT Structural and functional analysis of the nor-1 gene involved in the
 RT biosynthesis of aflatoxins by Aspergillus parasiticus.";
 RL Appl. Environ. Microbiol. 60:4078-4085(1994).
 CC -1- FUNCTION: MAY BE INVOLVED IN THE DEHYDRATION OF NOROLORINIC ACID
 CC TO FORM AFLATOXIN.
 CC -1- PATHWAY: AFLATOXIN BIOSYNTHESIS.
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 CC FAMILY (SDR). STRONG, TO E. NIDULANS STCE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: L27801; AAA58798.1; -
 CC PROSITE: PS00061; ADH_SHORT; FALSE_NEG.
 DR PFAM: PF00105; adh_short; 1.
 KW Oxidoreductase.
 FT ACT SITE 28 52 NAD OR NADP (BY SIMILARITY).
 FT ACT SITE 185 185 BY SIMILARITY.
 SQ SEQUENCE 271 AA; 29569 MW; A4798D11 CRC32;
 NORL_ASPPA Length: 271 February 14, 2000 08:02 Type: P Check: 7169 ..

1 MNGSLSDHQ ERLSTPYRDG PPEETVYLVY GASGIGRGL IEAFQRPFS
 51 TVVAMLRNRV TATPALSALT VARGSRMITY QLNDSERDA QAAYVQLARE
 101 HGVTHLDVVV ANAAMATNFG PASTMPLEHL QAHMAMVNTYA PVLLPQATRL

151 MLOOSKOAK FVLIGAPIST ITNMHDSYRA PLTAVGSKL AANYVWRKEH
 201 FENKWLIAFI IDPGHVQJDM GDQCARLMGR PQAPTVADS VAGICARIDE
 251 ATKETISGHH VHTDGSQLP W

11AA_SEQUENCE 1.0
 ID NIM_COTVA STANDARD; PRT; 72 AA.
 AC P24968;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 1 (EC 1.6.5.3) (FRAGMENT).
 GN MND1 OR NDI OR NADH1.
 OS coccinix coccinix japonica (Japanese quail).
 CC Mitochondrion.
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauiria; Aves;
 CC Neognathae; Galliformes; Phasianidae; Phasianinae; Columinix.
 CC [1]
 CC SEQUENCE FROM N.A.
 RP TISSUE=LIVER.
 RC MEDLINE; 91178819.
 RA DESJARDINS P., MORALS R.;
 RT "Nucleotide sequence and evolution of coding and noncoding regions of
 RT a quail mitochondrial genome.";
 RT J. Mol. Evol. 32:153-161(1991).
 CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE -> NAD(+) + UBIQUINOL.
 CC -1- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 1 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: X57246; CAA40522.1; -
 DR PIR: S25422; S25422.
 DR PROSITE: PS00667; COMPLEX1_NDI_1; PARTIAL.
 DR PROSITE: PS00668; COMPLEX1_NDI_2; PARTIAL.
 DR PFAM: PF00146; NADHdh; 1.
 KW Oxidoreductase; NAD; ubiquinone; Mitochondrion; Transmembrane.
 FT NON_TER 1
 SQ SEQUENCE 72 AA; 8286 MW; C1E65D31 CRC32;
 NIM_COTVA Length: 72 February 14, 2000 08:02 Type: P Check: 6471 ..

1 SFUSPSELF SITLAKVLL LSSSFLMIRA SYPRFRYDQL MHLWKNFLP
 51 LTLAKIMWT SMPYVAGLP PA

11AA_SEQUENCE 1.0
 ID NIM_ASNC STANDARD; PRT; 214 AA.
 AC Q1651;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5 (EC 1.6.5.3) (FRAGMENT).
 GN MND5 OR NDS OR NADH5.
 OS Anser caerulescens (Goose).
 CC Mitochondrion.
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauiria; Aves;
 CC Neognathae; Anseriformes; Anatidae; Anser.
 CC [1]
 CC SEQUENCE FROM N.A.
 RP MEDLINE; 94141942.
 RA QUINN T.W., WILSON A.C.;
 RT "Sequence evolution in and around the mitochondrial control region in
 RT birds.";
 RT J. Mol. Evol. 37:417-425(1993).
 CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE -> NAD(+) + UBIQUINOL.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: X75772; CA53393.1;
 CC Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
 CC
 CC FT NON TER 1 3006585 CRC32;
 CC SEQUENCE 214 AA: 23230 MW; 3C0E6A85 CRC32;
 NU6K_ANSC Length: 214 February 14, 2000 08:02 Type: P Check: 1655 ..

1 GSYDLOPSLN TSYLNTWALL LTMATAPFA TYSIRMTIIV CAGOTRIIPM
 51 VSMNENPLI TAPLTRALG STAGMITS FTPTKTPM TMLITRTAA
 101 ILMITIGIIL AELSNMHT LTPKPPLM NISSALGYEN PLVHFFSKN
 151 LLEKGNIAL HLDLSWLK MGPEGLAELQ VASKAATSM HTGLIKAYLG
 201 SFALSIYMI LMTM
 11AA_SEQUENCE 1.0 STANDARD: PRT: 167 AA.
 ID NU6M_APLI
 AC P34857;
 DT 01-FEB-1994 (Rel. 28, last sequence update)
 DT 15-DEC-1999 (Rel. 39, last annotation update)
 DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 6 (EC 1.6.5.3).
 GN ND6
 OS Apis mellifera ligustica (Common honeybee).
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;
 OC Apoidea; Apidae; Apis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=THORAX;
 RX MEDLINE: 93114603.
 RA CROZIER R.H., CROZIER Y.C.;
 RT "The mitochondrial genome of the honeybee Apis mellifera: complete
 RT sequence and genome organization";
 RL Genetics 133:97-117(1993).
 CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE -> NAD(+) + UBIQUINOL.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: L06178; AAB96808.1;
 CC Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
 CC
 CC FT NON TER 1 36864E1B CRC32;
 CC SEQUENCE 167 AA: 19971 MW; 36864E1B CRC32;
 NU6K_APLI Length: 167 February 14, 2000 08:02 Type: P Check: 6545 ..

1 MMLIIMLSK IFMSLSIMV LTIYLNIFN SPMLIYLI SYISIMSLM
 51 FTMSNMSL IMLIIVLS GMLIMSYFI SLINELKUK MKPIOTLFL
 101 IITMKIYNK LSONEYFNY FKNIDMLVY MKNLSLFLI MIMLIITLI
 151 LMTKIYIEK KTLRKKK
 11AA_SEQUENCE 1.0 STANDARD: PRT: 263 AA.
 ID NU6M_NEUCR

AC P40915;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, last sequence update)
 DT 15-JUL-1999 (Rel. 38, last annotation update)
 DE NADH-UBIQUINONE OXIDOREDUCTASE 24 KD SUBUNIT PRECURSOR (EC 1.6.5.3)
 DE (EC 1.6.99.3).
 GN NUO-24.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Euascomycetes; Pyrenomycetes;
 OC Sordariatales; Sordariaceae; Neurospora.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 95034970.
 RA AZEVEDO J.E., DUARTE M., BELO J.A., WERNER S., VIDEIRA A.;
 RT "Complementary DNA sequences of the 24 kDa and 21 kDa subunits of
 RT complex I from Neurospora".
 RL Biochim. Biophys. Acta 1168:159-161(1994).
 CC -1- FUNCTION: TRANSFER OF ELECTRONS FROM NADH TO THE RESPIRATORY
 CC CHAIN. THE IMMEDIATE ELECTRON ACCEPTOR FOR THE ENZYME IS BELIEVED
 CC TO BE UBIQUINONE. COMPONENT OF THE FLAVOPROTEIN-SULFUR (FP)
 CC FRAGMENT OF THE ENZYME.
 CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE -> NAD(+) + UBIQUINOL.
 CC -1- COFACTOR: 2FE-2S IRON-SULFUR CLUSTER N 1B.
 CC -1- SUBUNIT: COMPLEX I IS COMPOSED OF ABOUT 40 DIFFERENT SUBUNITS.
 CC THIS IS A COMPONENT OF THE FLAVOPROTEIN FRACTION.
 CC -1- SUBCELLULAR LOCATION: MATRIX AND CYTOPLASMIC SIDE OF THE
 CC MITOCHONDRIAL INNER MEMBRANE.
 CC -1- SIMILARITY: BELONGS TO THE COMPLEX I 24 KD SUBUNIT FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: X78083; CA54990.1;
 CC PROSITE: PS01099; COMPLEX1_24K: 1.
 DR PFWA; PF01257; COMPLEX1_24KD: 1.
 KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transit peptide;
 KW Iron-sulfur.
 FT TRANSIT 1 42
 FT CHAIN 43 263
 FT METAL 143 143
 FT METAL 148 148
 FT METAL 186 186
 FT METAL 190 190
 FT SEQUENCE 263 AA: 28842 MW; 0C33B3CC CRC32;
 NU6M_NEUCR Length: 263 February 14, 2000 08:02 Type: P Check: 3961 ..

1 MATKILPLM RIVAKRATRL STRPSTIAPV SRACISISAR RSDIPLMHR
 51 NTPDNNPPIP FKPSADNEKY IEIITKRPP QYKKAAMPL LDGROHGF
 101 GSISVMEVA RLIEPMPRY VEVASFYMY NRTPVGRFHV QACTTPCOL
 151 GGCGSDVIYK AIKEHGIQK GETTPDGLFT FIEVEICGAC ANAPVQIND
 201 DYFEDLPET IKOVLALKE SVTDVSKAPO PGROSGROT ENAAGLSLTL
 251 SEYGPETVR SDL
 11AA_SEQUENCE 1.0 STANDARD: PRT: 215 AA.
 ID NU6M_BRAOL
 AC P42027;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, last sequence update)
 DT 15-JUL-1999 (Rel. 38, last annotation update)
 DE NADH-UBIQUINONE OXIDOREDUCTASE 20 KD SUBUNIT PRECURSOR (EC 1.6.5.3)
 DE (EC 1.6.99.3) (COMPLEX I-20KD) (CI-20KD).

OS Brassica oleracea (Cauliflower).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC eumhyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 CC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 CC Brassica.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. SHOGUN; TISSUE-FLORETS;
 RX MEDLINE: 95334517.
 RA POGSON B.J., DOMNS C.G., DAVIES K.M., MORRIS S.C.,
 RA BUCHANAN-WOLLASTON V.,
 RT "Nucleotide sequence of a cDNA clone from broccoli with high identity
 RT with the PEST subunit of NADH:ubiquinone oxidoreductase";
 RL Plant Physiol. 108:859-860(1995).
 CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
 CC -1- COFACTOR: MAY CONTAIN AN IRON-SULFUR 4FE-4S CLUSTER.
 CC -1- SUBUNIT: COMPLEX I IS COMPOSED OF ABOUT 40 DIFFERENT SUBUNITS.
 CC -1- SIMILARITY: BELONGS TO THE COMPLEX I 20 KD SUBUNIT FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X82274; CAA57725.1; -
 CC DR PROSITE; PS01150; COMPLEX1_20K; 1.
 CC DR PFMW; PF01058; oxidoreduc_46; 1.
 CC KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transist peptide;
 CC Iron-sulfur; 4Fe-4S;
 CC FT TRANSIT ? 215
 CC FT CHAIN 1
 CC FT NADH-UBIQUINONE OXIDOREDUCTASE 20 KD
 CC FT SUBUNIT.
 CC FT IRON-SULFUR (4FE-4S) (POTENTIAL).
 CC FT IRON-SULFUR (4FE-4S) (POTENTIAL).
 CC FT IRON-SULFUR (4FE-4S) (POTENTIAL).
 CC FT IRON-SULFUR (4FE-4S) (POTENTIAL).
 CC FT METAL 185 185
 CC FT METAL 185 185
 CC FT SEQUENCE 215 AA; 23839 MW; DDDA728B CRC32;
 CC SQ
 CC NUKM_BROOL Length: 215 February 14, 2000 08:02 Type: P Check: 9647 ..
 CC 1 MAMTRNTAT RLPVLVQSHR AAVSHLTS LPALSPATTP TSYTPGPPS
 CC 51 TSAPPGSLK TAEFVISKVD DLMKARKGS IMPATISLAC CAVENMHTGA
 CC 101 ARYDLDFGI IFRSPROSD CMYAGILTN KMAPALRKVT DDMPEPRWT
 CC 151 SMGSCANGG YHYSTSVVR GCDRIVPVDI YVDCPPTAE ALLYGLLOLO
 CC 201 KRINRKDFL HMMNR
 CC 11AA_SEQUENCE 1.0
 CC ID NUOJ_ECOLI STANDARD; PRT; 184 AA.
 CC AC P33605; P78236;
 CC DT 01-FEB-1994 (Rel. 28, Created)
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
 CC DT 01-NOV-1997 (Rel. 35, Last annotation update)
 CC DE NADH DEHYDROGENASE I CHAIN J (EC 1.6.5.3) (NADH-UBIQUINONE
 CC OXIDOREDUCTASE CHAIN 10) (NUO10).
 CC GN NUOJ
 CC OS Escherichia coli.
 CC CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN-K12 / AN387;
 CC RX MEDLINE: 93369724.
 CC RA WEDNER U., GEIER S., FLOCK A., FRIEDRICH T., LEIF H., WEISS H.,
 CC "The gene locus of the proton-translocating NADH: ubiquinone
 CC oxidoreductase in Escherichia coli. Organization of the 14 genes and
 CC RT relationship between the derived proteins and subunits of

RT mitochondrial complex I";
 RL J Mol. Biol. 233:109-122(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE: 97426617.
 RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
 RA KILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
 RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
 RA MAU B., SHAO Y.,
 RT "The complete genome sequence of Escherichia coli K-12";
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RA ALBA H., BABA T., FUJITA K., HAYASHI K., HONJO A., HORIUCHI T.,
 RA IKEMOTO K., INADA T., ISONO K., ITOH T., KANAI K., KASAI H.,
 RA KASHIMOTO K., KIM S., KIMURA S., KITAGAWA M., KITAKAWA M., MAKINO K.,
 RA MASUDA S., MIKI T., MIZOBUCHI K., MORI H., MOTOMURA K., NAKAMURA Y.,
 RA NASHIMOTO H., NISHIO Y., OSHIMA T., SATO N., SAMPEI G., SEKI Y.,
 RA TAGAMI H., TAKEMOTO K., MADA C., YAMAMOTO Y., YANO M.,
 RA Submitted (Jan-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: TRANSFER OF ELECTRONS FROM NADH TO THE RESPIRATORY
 CC CHAIN. THE IMMEDIATE ELECTRON ACCEPTOR FOR THE ENZYME IS BELIEVED
 CC TO BE UBIQUINONE. DOES COUPLE THE REDOX REACTION TO PROTON
 CC TRANSLLOCATION AND THUS CONSERVES THE REDOX ENERGY IN A PROTON
 CC GRADIENT.
 CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
 CC -1- SUBUNIT: COMPOSED OF 14 DIFFERENT SUBUNITS. SUBUNITS NUO4, H, J,
 CC K, L, M, N CONSTITUTE THE MEMBRANE SECTOR OF THE COMPLEX.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: TO POLYPEPTIDE 6 OF THE NADH-UBIQUINOL OXIDOREDUCTASE
 CC OF CHLOROPLASTS OR MITOCHONDRIA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X68301; CAA48369.1; -
 CC DR EMBL: AE000317; AAC75340.1; -
 CC DR EMBL: D90859; CAB22037.1; -
 CC DR PIR: S37067; S37067.
 CC DR PIR: S38319; S38319.
 CC DR ECGENE: EGI2090; NUOJ
 CC DR PFMW: PF00499; oxidoreduc_43; 1.
 CC KW Oxidoreductase; NAD; Ubiquinone; Transmembrane.
 CC FT TRANSMEM 1 21
 CC FT TRANSMEM 28 74
 CC FT TRANSMEM 54 74
 CC FT TRANSMEM 92 112
 CC FT TRANSMEM 138 158
 CC FT TRANSMEM 85 85
 CC FT CONFLICT 126 126
 CC FT CONFLICT 126 126
 CC FT SEQUENCE 184 AA; 19874 MW; 29FD512 CRC32;
 CC SQ
 CC NUOJ_ECOLI Length: 184 February 14, 2000 08:02 Type: P Check: 2909 ..
 CC 1 MEFAFYICGL IAILATLTVI THNPVHAL YLISLTAIS GVFFSLGAVF
 CC 51 AGALEITIVA GAIVLFEV VMMLNLGSGE IGEKQWLKP QWIGPAILS
 CC 101 AMLVYIIVA ILGVNDGID GTPISAKAVG IILFGYVLA VELASMLLA
 CC 151 GIVAFIVGR EERAGEVLSN RKDSARAKT ESHA
 CC 11AA_SEQUENCE 1.0
 CC ID NUOJ_ECOLI STANDARD; PRT; 201 AA.
 CC AC P19968;
 CC DT 01-FEB-1991 (Rel. 17, Created)

DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE MADH-UBIQUINONE OXIDOREDUCTASE 21.3 KD SUBUNIT (EC 1.6.5.3)
 DE (EC 1.6.99.3).
 OS Neurospora crassa.
 CC Eukaryota; Fungi; Ascomycota; Euascomycetes; Pyrenomycetes;
 CC Sordariatales; Sordariaceae; Neurospora.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 90147712.
 RA VIEIRA A., TROPSCHUG M., WERNER S.;
 RT "Primary structure, in vitro expression and import into mitochondria
 of a 29/21-kDa subunit of complex I from Neurospora crassa.";
 RL Biochem. Biophys. Res. Commun. 166:280-285(1990).
 CC -1- FUNCTION: TRANSFER OF ELECTRON FROM NADH TO THE RESPIRATORY
 CHAIN. THE IMMEDIATE ELECTRON ACCEPTOR FOR THE ENZYME IS BELIEVED
 TO BE UBIQUINONE.
 CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE -> NAD(+) + UBIQUINOL.
 CC -1- SUBUNIT COMPLEX I IS COMPOSED OF ABOUT 30 DIFFERENT SUBUNITS.
 CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE. IMPORTATION
 INTO MITOCHONDRION IS MEMBRANE POTENTIAL DEPENDENT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M32244; AAA33570.1; -
 DR PIR: A34051; A34051.
 DR OXIDOREDUCTASE; NAD; Ubiquinone; Mitochondrion; Membrane.
 SQ SEQUENCE 201 AA; 21347 MW; 32F6D36F CRC32;
 NUZM_NEUCR Length: 201 February 14, 2000 08:02 Type: P Check: 7750 ...
 1 MASKVYGVY KTAGGVVY SQKTVSYG VMERIRRAFA IDPRKSGV
 51 LVYNNPSP GSLDPLAYDD PYTPADIA DNPYKRRAR RNYRSLVYG
 101 QAQAVALLSY GSATHPRVEL VGENGSKOLV AAOEGKGTGG LAKTFESTGV
 151 EAKVLVLAET GGLPPLPGE KLGEGKWDV YVYQLAEEPS YSEAYPCRSF
 201 S
 !!AA_SEQUENCE 1.0 STANDARD: PRT; 306 AA.
 ID OPB_ECOLI
 AC P31132; P76026; P77550;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE OLIGOPETIDE TRANSPORT SYSTEM PERMEASE PROTEIN OPB.
 DE OLIGOPETIDE TRANSPORT SYSTEM PERMEASE PROTEIN OPB.
 GN OPB.
 OS Escherichia coli.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE: 97426617.
 RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
 RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
 RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
 RA MAU B., SHAO Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RA AIBA H., BABA T., FUJITA K., HAYASHI K., HONJO A., HORIUCHI T.,
 RA IKEMOTO K., INADA T., ISONO K., ISONO S., ITOH T., KANMI K., KASAI H.,

RA KASHIMOTO K., KIM S., KIMURA S., KITAGAWA M., KITAKAWA M., MAKINO K.,
 RA MASUDA S., MIKI T., MIOBUCHI K., MORI H., MOTOKURA K., NAKAMURA Y.,
 RA NASHIMOTO H., NISHIO Y., OSHIMA T., SATO N., SAMPEI G., SEKI Y.,
 RA TAGAMI H., TAKEMOTO K., WADA C., YAMAMOTO Y., YANO M.;
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-16 FROM N.A.
 RX MEDLINE: 90256748.
 RA KASHIMOTO K., TAGUCHI Y., SAKAI Y., KOBAYASHI H., IGARASHI K.;
 RT Identification of the polyamine-induced protein as a periplasmic
 oligopeptide binding protein.
 RL J. Biol. Chem. 265:8387-8391(1990).
 CC -1- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
 FOR OLIGOPETIDES; PROBABLY RESPONSIBLE FOR THE TRANSLLOCATION OF
 THE SUBSTRATE ACROSS THE MEMBRANE.
 CC -1- SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-
 PROTEIN-DEPENDENT TRANSPORT SYSTEMS. BELONGS TO THE OPBc
 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE000223; AAC74326.1; -
 DR EMBL: D90763; CAB20709.1; -
 DR EMBL: D90852; CAB21967.1; -
 DR EMBL: J05433; -; NOT_ANNOTATED_CDS.
 DR PIR: B36263; B36263.
 DR ECOCENF: PG10675; OPB.
 DR PROSITE: PS00402; BPD_TRANSF. 1.
 DR PIRAF: PF00528; BPD_TRANSF. 1.
 KM transport; Peptide transport; Transmembrane; Inner membrane.
 FT DOMAIN 1
 FT TRANSMEM 9
 FT DOMAIN 1
 FT TRANSMEM 10
 FT TRANSMEM 30
 FT TRANSMEM 99
 FT TRANSMEM 100
 FT TRANSMEM 122
 FT TRANSMEM 137
 FT TRANSMEM 158
 FT TRANSMEM 172
 FT TRANSMEM 173
 FT TRANSMEM 190
 FT TRANSMEM 191
 FT TRANSMEM 227
 FT TRANSMEM 250
 FT TRANSMEM 251
 FT TRANSMEM 272
 FT TRANSMEM 293
 FT TRANSMEM 294
 FT DOMAIN 306
 SQ SEQUENCE 306 AA; 33443 MW; 3FCC09F5 CRC32;
 OPB_ECOLI Length: 306 February 14, 2000 08:02 Type: P Check: 5147 ..
 1 MKFTLRCL EALPILFILI TISFMRRA PGSPFPGENT LPPEVMANIE
 51 AKYHNPIM TOYFTSLKOL AHGPGPSFK YKDYSVNDLV ASSPVSATKL
 101 GAAPFLAVI LGVSAGVIA LKONTKWDYT VNGLAMTVV IPSFVAPPL
 151 VMFAILLHW LPEGGNGGA LKEMILPWA LSLAVIASIA RITRGSMEIV
 201 LHSNFRTAR AKGLPRRLI LRHALKPLAL PVLSYNGPAF VGIITGSMTI
 251 ETIVGLPGIG QLVNGALNR DYSLVSLTI LVGALTIIFN AIVDVLAVI
 301 DPKIRY
 !!AA_SEQUENCE 1.0 STANDARD: PRT; 306 AA.
 ID OPB_SALT
 AC P08005;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)

01-FEB-1995 (Rel. 31, last annotation update)
 DE OLIGOPEPTIDE TRANSPORT SYSTEM PERMEASE PROTEIN OPB.
 GN OPB.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-LT2.
 RX MEDLINE: 86011222.
 RA HILES I.D., GALLAGHER M.P., JAMESON D.J., HIGGINS C.F.:
 RT "Molecular characterization of the oligopeptide permease of
 J. Mol. Biol. 195:125-142(1987)."
 RL (1)
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-LT2.
 RX MEDLINE: 92149312.
 RA PEARCE S.R., MIMMACK M.L., GALLAGHER M.P., GILEADI U., HYDE S.C.,
 RT HIGGINS C.F.:
 "Membrane topology of the integral membrane components, OPB and
 OPB, of the oligopeptide permease of *Salmonella typhimurium*."
 RL Mol. Microbiol. 6:47-57(1992).
 CC -1- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
 FOR OLIGOPEPTIDES. PROBABLY RESPONSIBLE FOR THE TRANSLLOCATION OF
 THE SUBSTRATE ACROSS THE MEMBRANE.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
 CC -1- SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-
 PROTEIN-DEPENDENT TRANSPORT SYSTEMS. BELONGS TO THE OPB/C
 SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X05491; CAA29040.1; -
 DR PIR: B29333; OREBOB.
 DR SYNGENE: SG10268; OPB.
 DR PROSITE: PS00402; BPD_TRANSF_INN_MEMBER; 1.
 DR PIR: PF00528; BPD_transp; 1.
 KW Transport; Peptide transport; Transmembrane; Inner membrane.
 FT DOMAIN 1 9
 FT TRANSMEM 30
 FT DOMAIN 31 99
 FT TRANSMEM 100 121
 FT TRANSMEM 122 137
 FT TRANSMEM 138 158
 FT TRANSMEM 159 172
 FT TRANSMEM 173 190
 FT TRANSMEM 191 226
 FT TRANSMEM 227 250
 FT TRANSMEM 251 271
 FT TRANSMEM 272 293
 FT TRANSMEM 294 306
 FT DOMAIN 306 AA: 33420 MW: 8A020E68 CRC32;
 SQ SEQUENCE

OPB_SALTY Length: 306 February 14, 2000 08:02 Type: P Check: 4920

1 MLKFLIRCL EAIPLFLILI TISFFMRLA PGSPFTGERA LPPEVLNIE
 51 AKYHLNDPIM TOYFSYKOL AHGDFGSPK YKDYVNDVY AASEFVSAKL
 101 GAANFLAVI IGSAGVIA LKONTWDTY VMGFATGVV IPSEFVAPL
 151 VMVFAITLW LPPGGWNGA LKFWILPVA LSLAYIASIA RITRSMIEV
 201 LHSNFIETAR AKGLPWRII FRHALKPLL PVLSTYGPVF VGIIGSNVI
 251 ETIYGLPGIG QLFVNGALNR DYSVLVSLTI LVGALITLEN ALVDYLVAVI

301 DEKIRY
 !!AA_SEQUENCE 1.0
 ID OPS1_CALVI STANDARD; PRT; 371 AA.
 AC P22269;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 01-DEC-1992 (Rel. 24, Last annotation update)
 DE OPS1_CALVI (OUTER R1-R6 PHOTORECEPTOR CELLS OPSIN).
 GN RH1 OR NINAE
 OS Calliphora vicina (Blue blowfly) (Calliphora erythrocephala).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Felygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Oestridae; Calliphoridae; Calliphora.
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 91009258.
 RA HUBER A., SMITH D.P., ZUKER C.S., PAULSEN R.:
 RT "Opsin of Calliphora peripheral photoreceptors R1-6. Homology with
 Drosophila Rh1 and posttranslational processing.";
 RL J. Biol. Chem. 265:17906-17910(1990).
 CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
 MEDIANE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
 LINKED TO CIS-RETINAL.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
 BE PHOSPHORYLATED.
 CC -1- MISCELLANEOUS: EACH EYE IS COMPOSED OF 800 FACETS OR OMMATIDIA.
 EACH OMMATIDIUM CONTAINS 8 PHOTORECEPTOR CELLS (R1-R8), THE R1 TO
 R6 CELLS ARE OUTER CELLS, WHILE R7 AND R8 ARE INNER CELLS.
 CC -1- MISCELLANEOUS: OPSIN RH1 HAS AN ABSORPTION MAXIMA AT 480 NM.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC OPSIN SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M58334; AAA62725.1; -
 DR PIR: A39234; A39234.
 DR GCRDB: GCR_0009; -
 DR PROSITE: PS00237; G-PROTEIN_RECEPTOR; 1.
 DR PROSITE: PS00238; OPSIN; 1.
 DR PIR: PF00001; 7tm.1; 1.
 KW Photoreceptor; Retinal protein; Transmembrane; Phosphorylation;
 KW Glycoprotein; G-protein coupled receptor; Vision.
 FT DOMAIN 1 48
 FT TRANSMEM 72
 FT DOMAIN 73 84
 FT TRANSMEM 85 110
 FT DOMAIN 111 124
 FT TRANSMEM 125 144
 FT DOMAIN 145 163
 FT TRANSMEM 164 187
 FT DOMAIN 188 211
 FT TRANSMEM 212 239
 FT DOMAIN 240 274
 FT TRANSMEM 275 298
 FT DOMAIN 299 305
 FT TRANSMEM 306 330
 FT DOMAIN 331 371
 FT TRANSMEM 371 388
 FT CARBOHYD 18
 FT DISULFID 121
 FT BINDING 317
 FT SEQUENCE 371 AA: 41289 MW: 6D98882F CRC32;
 SQ

OPS1_CALVI Length: 371 February 14, 2000 08:02 Type: P Check: 9791

1 MERYSTPLIG PSFALTNIGS VTDKVTDPMA HLVDHYNQF PAHEPWAKF

```

51 LAAYWLVLT ISMGNGVYI YIFSTTKSLR TPANLVNL AISDFGIMT
101 NTPMGINLF YETWYGLPLM CDYIGGLGSA FGCSSILSMC MISLDRYNI
151 VYGMAGQPMY IKLAIMKIAL IMFASIMWL APVQMSRYV PEGNLSCGI
201 DYLERDMNDR SYLIFYSIFV YLPLFLICY SYWFLAIVS AHEKAREQA
251 KKMANKSLRS SEDADKSAEG KLAVALVNI SLMEFAMPY TIINTGLFR
301 YGLPLPLNTI WGACFAKSA CYNPIYGIS HPKVIALKE KPCPCVGVY
351 DGGKASDATS QATNNESETK A

11AA_SEQUENCE 1.0
ID OPS1_DROME STANDARD: PRT: 373 AA.
AC P06002;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE OPSIN RH1 (OUTER R1-R6 PHOTORECEPTOR CELLS OPSIN).
GN RH1 OR NINAE.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 85176937.
RA O'LOUSA J.E., BAEHR W., MARTIN R.L., HIRSH J., PAK W.L.,
RA APPLEBURY M.L.;
RT "The Drosophila rh1a gene encodes an opsin.";
RL Cell 40:839-850(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 85176938.
RA ZUKER C.S., COMAN A.F., RUBIN G.M.;
RT "Isolation and structure of a rhodopsin gene from D. melanogaster.";
RL Cell 40:851-858(1985).
CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
CC MEDIANE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
CC LINKED TO CIS-RETINAL.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
CC BE PHOSPHORYLATED.
CC -1- MISCELLANEOUS: EACH DROSOPHILA EYE IS COMPOSED OF 800 FACETS OR
CC OMATIDIA. EACH OMATIDIUM CONTAINS 8 PHOTORECEPTOR CELLS (R1-R6),
CC THE R1 TO R6 CELLS ARE OUTER CELLS, WHILE R7 AND R8 ARE INNER
CC CELLS.
CC -1- MISCELLANEOUS: OPSIN RH1 HAS AN ABSORPTION MAXIMA AT 480 NM.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC OPSIN SUBFAMILY.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K02315; AAA28733.1; -;
DR EMBL; K02320; AAA28735.1; ALT-SEQ.
DR EMBL; K02316; AAA28735.1; JOINED.
DR EMBL; K02317; AAA28735.1; JOINED.
DR EMBL; K02318; AAA28735.1; JOINED.
DR EMBL; K02319; AAA28735.1; JOINED.
DR PIR; A22012; COFE.
DR GCRDB; GCR 0028; -;
DR GCRDB; GCR 0029; -;
DR FLYBASE; FBgn0002940; rh1a.
DR PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.

```

```

DR PROSITE; PS00238; OPSIN; 1.
DR PFAM; PFO0001; 7tm_1; 1.
KW Photoreceptor; Retinal protein; Transmembrane; Phosphorylation;
KW Glycoprotein; G-protein coupled receptor; Vision.
FT DOMAIN 1 49
FT TRANSMEM 50 74
FT DOMAIN 75 86
FT TRANSMEM 87 112
FT DOMAIN 113 126
FT TRANSMEM 127 146
FT DOMAIN 147 165
FT TRANSMEM 166 189
FT DOMAIN 190 213
FT TRANSMEM 214 241
FT DOMAIN 242 276
FT TRANSMEM 277 300
FT DOMAIN 301 307
FT TRANSMEM 308 332
FT DOMAIN 333 373
FT CARBOHYD 20 20
FT DISULFID 123 200
FT BINDING 319 319
SQ SEQUENCE 373 AA; 41494 MW; 43B10BFA CRC32;

OPS1_DROME Length: 373 February 14, 2000 08:02 Type: P Check: 4358
1 MESFAVAQAQ LGPHFAPLSN GSVVDKVTPTD MAHLSPYWN QEPAMDPIWA
51 KILFAYIMI GMSMGNGV YIFATTKS LTPANLVNI NLATSDFGIM
101 ITNTPMGIN LFETWYGLP MMCDIYAGLG SAFGCSISMS MGMTSLDRYQ
151 VYGMAGQPMY KTIPLALSKI AYIMFSSIM CLAPFGNSR YVPGNLTSC
201 GIDLERDMN PRSYLIFYSI FVYIPLFL CYSYFIINA VSAHEKAMRE
251 QAKMANKSL RSEDEAKSA EKLAKVALV TITLWMAMT PYLVINCMGL
301 FREFGLPLN TIWGAFAKSA AACYNPIYVG ISHPRYRLAL KKPCCVGVG
351 KVDDKSSDA QSQATASEAE SKA

11AA_SEQUENCE 1.0
ID OPS1_DROME STANDARD: PRT: 374 AA.
AC P28678;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE OPSIN RH1 (OUTER R1-R6 PHOTORECEPTOR CELLS OPSIN).
GN RH1 OR NINAE.
OS Drosophila pseudoobscura (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93012921.
RA CARULLI J.P., HARVEY D.L.;
RT "Variable rates of evolution among Drosophila opsin genes.";
RL Genetics 133:193-204(1992).
CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
CC MEDIANE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
CC LINKED TO CIS-RETINAL.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
CC BE PHOSPHORYLATED.
CC -1- MISCELLANEOUS: EACH DROSOPHILA EYE IS COMPOSED OF 800 FACETS OR
CC OMATIDIA. EACH OMATIDIUM CONTAINS 8 PHOTORECEPTOR CELLS (R1-R6),
CC THE R1 TO R6 CELLS ARE OUTER CELLS, WHILE R7 AND R8 ARE INNER
CC CELLS.
CC -1- MISCELLANEOUS: OPSIN RH1 HAS AN ABSORPTION MAXIMA AT 480 NM.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

```

OPSIN SUBFAMILY

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X65877; CAA46708.1; -
 DR PIR: S24605; S24605.
 DR GCRDB: GCR_0429; -
 DR GCRDB: GCR_0509; -
 DR FLYBASE: FBgn0012733; Dpse\unae.
 DR PROSITE: PS00237; G_PROTEIN_RECEPTOR; 1.
 DR PROSITE: PS00238; OPSIN; 1.
 DR PFM: PFO0001; 7tm_1; 1.
 DR Photoreceptor; Retinal protein; Transmembrane; Phosphorylation;
 KW Glycoprotein; G-protein coupled receptor; Vision.
 FT DOMAIN 1 49 EXTRACELLULAR
 FT TRANSMEM 50 74 CYTOPLASMIC
 FT TRANSMEM 75 86 2 (POTENTIAL)
 FT TRANSMEM 87 112 EXTRACELLULAR
 FT TRANSMEM 113 126 3 (POTENTIAL)
 FT TRANSMEM 127 146 CYTOPLASMIC
 FT TRANSMEM 147 165 4 (POTENTIAL)
 FT TRANSMEM 166 190 EXTRACELLULAR
 FT TRANSMEM 191 214 5 (POTENTIAL)
 FT TRANSMEM 215 242 CYTOPLASMIC
 FT TRANSMEM 243 277 6 (POTENTIAL)
 FT TRANSMEM 278 301 EXTRACELLULAR
 FT TRANSMEM 302 308 7 (POTENTIAL)
 FT TRANSMEM 309 333 CYTOPLASMIC
 FT TRANSMEM 334 374 PROBABLE
 FT CARBOHYD 20 20 POTENTIAL
 FT DISULFID 123 201 RETINAL CHROMOPHORE
 FT BINDING 320 320
 FT SEQUENCE 374 AA; 41604 MW; AAC8CB5 CRC32;
 OS1_Drops Length: 374 February 14, 2000 08:02 Type: P Check: 7241 ..
 1 MOSFAVATQ LGPOFAPSN GSVDKVTDP MAHLISPYMD QPAMPPIVA
 51 KLTAVMIIT GMISMGNGV VIYIPATKS LRTANLIVI NLAISDFGIN
 101 INTFMNGIN LYFETWLGK MMCDIYAGIG SAFGCCSIWS MCMISIDRQ
 151 VYVKMAGRP MTPLALGKI AYIWMSTIW CCLAPYVQWS RYVPEGNLIS
 201 CGIDYLERDW NPSYLIFFS IYVYIPLFL ICYSWFLIA AVAHEKAR
 251 EOKAKMNVKS LRSESDADKS AEGKLAKVAL VTISLWPMW TPLVINCWG
 301 LKFEGLTPL NTIWACFAK SAACYNPIVY GISHPKRYLA LKEKPCCVF
 351 GKYDDKSSSE AOSQATTSEA ESKA
 !!AA_SEQUENCE 1.0
 ID OTCA_VIBS2 STANDARD; PRT; 301 AA.
 AC P96172;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ORNITHINE CARBAMOYLTRANSFERASE, ANABOLIC (EC 2.1.3.3) (OTCASE).
 GN ARGF.
 OS Vibrio sp. (strain 2693).
 CC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA LIANG Z.Y., DEMAREZ M., LECRAIN C., BAETENS M., ROOYERS M.,
 RA GLASDORF N.;
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.

 CC -1- CATALYTIC ACTIVITY: CARBAMOYL PHOSPHATE + ORNITHINE - CITRULLINE
 CC + ORTHOPHOSPHATE.
 CC -1- PATHWAY: ARGinine BIOSYNTHESIS.
 CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: BELONGS TO THE ATCASES/OTCASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Y11033; CAA11921.1; -
 DR HSSP: P04391; 1AKM.
 DR PROSITE: PS00097; CARBAMOYLTRANSFERASE; 1.
 DR PFM: PFO0185; OTCase; 1.
 KW Arginine biosynthesis; Transferase.
 FT SEQUENCE 301 AA; 32825 MW; 3CA3E90D CRC32;
 OTCA_VIBS2 Length: 301 February 14, 2000 08:02 Type: P Check: 7498 ..
 1 MENLSYKDL SKQIIDLDA LAKAVKANPA EYSQALAGS IYIYERSL
 51 RRTVPFGIG HKIGSHAVYL DAONGAIGER EYKDFRANI SRWADAIVAR
 101 VVSHKTEGL VEGSVYPVYN SLCDLYHPCQ ALADFLTISE HYEDVSKYKL
 151 AYVEGNNVT HSLMTGAIL GAEYVACPR GSSPDQIYK QAMALAEISG
 201 KRIYVNDLD AYVDYVITG DTWVSMGDDT PLAQVEKRYA PQINKALLM
 251 RTGIKHLVLC QPAHRELEIT SEVMDGHSIL IEDQENRHH NONAVLITLL
 301 K
 !!AA_SEQUENCE 1.0
 ID OTCA_SCHPO STANDARD; PRT; 327 AA.
 AC P31317;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ORNITHINE CARBAMOYLTRANSFERASE PRECURSOR (EC 2.1.3.3) (OTCASE)
 GN ARG3 OR SPAC49.10.
 OS Schizosaccharomyces pombe (fission yeast).
 CC Schizosaccharomycetes; Eukaryota; Fungi; Ascomycota; Archiaascomycetes;
 CC Schizosaccharomycetales; Schizosaccharomycetaceae;
 CC Schizosaccharomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN-975;
 RC MEDLINE: 92209520.
 RA VAN HOFFEL C., DOBOIS E., MESSENGUY F.;
 RT Cloning and sequencing of arg3 and arg11 genes of
 RT Schizosaccharomycetes pombe on a 10-kb DNA fragment. Heterologous
 RT expression and mitochondrial targeting of their translation
 RT products.
 RL Eur. J. Biochem. 205:33-43(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA STRAIN-972;
 RC CONNOR R., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: CARBAMOYL PHOSPHATE + ORNITHINE - CITRULLINE
 CC + ORTHOPHOSPHATE.
 CC -1- PATHWAY: FIRST STEP IN ARGinine BIOSYNTHESIS.
 CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
 CC -1- SIMILARITY: BELONGS TO THE ATCASES/OTCASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: X63577; CAA45133.1; ..
 DR EMBL: Z69727; CAA93560.1; ..
 DR PIR: S22901; OMDP.
 DR HSP: O51742; IALS.
 DR PROSITE: PS00097; CARBAMOYLTRANSFERASE: 1.
 DR PFAM: PF00185; OTCace; 1.
 KW Arginine biosynthesis; Transferase; Mitochondrion; Transit peptide.
 FT TRANSIT 1 327 MITOCHONDRION.
 FT CHAIN 1 327 ORNITHINE CARBAMOYLTRANSFERASE.
 SQ SEQUENCE 327 AA; 36119 MW; 9CC43E5C CRC32;
 OTC_SCHPO Length: 327 February 14, 2000 08:02 Type: P Check: 9240 ..

1 MSFKFPRHL LSIRDSRGE IVKLIDRSSE IKOAYKONFQ NRRSYOMSGL
 51 SSCWAMIFS KRSTRTRVSV ESAYSCLSGN AMFLGKDIDQ LGVNESTYDT
 101 SKVSSWVSG IYAVNKRYSO VATLAKHASC PVINGLODTE HPLQALADLL
 151 TIKETFSFD GLKVAVWGDA NNVLHDLMLA NAKVGIHTSV AKPKOVNVRD
 201 DILSIWNEA NENGSTFEIV NDPVAVAKNA DIVVTDWIS MGQAEKREQR
 251 LKQFTGQVY GEIMKIAKPS CKFMHCLPRH PEEVSDEVY GENSLVPEEA
 301 ENKRTIVAV LEALVNRGE ILPPASA

11AA_SEQUENCE 1.0 STANDARD; PRT: 341 AA.
 ID OXDD_BOVIN 002846; Created
 AC P1228; 002846; (Rel. 26, Last sequence update)
 DT 01-JUL-1993 (Rel. 38, Last annotation update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE D-ASPARTATE OXIDASE (EC 1.4.3.1) (DASOX) (DDO).
 GN DDO.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
 OC Bovinae; Bos.
 RN [1]
 RP SEQUENCE OF 1-338.
 RC TISSUE-KIDNEY;
 RX MEDLINE: 92291057.
 RA NEGRI A., CECILIANI F., TEDESCHI G., SIMONIC T., RONCHI S.;
 RT "The primary structure of the flavoprotein D-aspartate oxidase from
 RT beef kidney".
 RT J. Biol. Chem. 267:11865-11871(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-KIDNEY CORTEX;
 RX MEDLINE: 97220379.
 RA SIMONIC T., DUGA S., NEGRI A., TEDESCHI G., MALCOVANTI M.,
 RA TECHINI M.L., RONCHI S.;
 RT cDNA cloning and expression of the flavoprotein D-aspartate oxidase
 RT from bovine kidney cortex.
 RT Biochem. J. 322:729-735(1997).
 RL -1- CATALYTIC ACTIVITY: D-ASPARTATE + H(2)O + O(2) = OXALACETATE +
 CC NH(3) + H(2)O(2).
 CC -1- COFACTOR: FAD OR 6-HYDROXYFLAVIN ADENINE DINUCLEOTIDE.
 CC -1- SUBUNIT: MONOMER.
 CC -1- SUBCELLULAR LOCATION: PEROXISOMAL.
 CC -1- SIMILARITY: BELONGS TO THE DAMOX/DASOX FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: X65310; CAA64622.1; ..
 DR PIR: A44332; A44132.
 DR HSP: P00371; IDAO.
 DR PROSITE: PS00677; DAO; 1.
 DR PFAM: PF01265; DAO; 1.
 KW Oxidoreductase; Flavoprotein; FAD; Peroxisome.
 FT MOD_RES 1 20 BLOCKED.
 FT NE_BIND 6 20 FAD (ADP PART) (POTENTIAL).
 FT ACT_SITE 223 223 BY SIMILARITY.
 FT ACT_SITE 302 302 BY SIMILARITY.
 FT SITE 339 341 MICROBODY TARGETING SIGNAL (POTENTIAL).
 FT VARIANT 228 228 V -> I (IN SOME MOLECULES).
 FT CONFLICT 274 274 K -> R (IN REF. 1).
 FT CONFLICT 283 283 S -> G (IN REF. 1).
 SQ SEQUENCE 341 AA; 37659 MW; 842B960C CRC32;
 OXDD_BOVIN Length: 341 February 14, 2000 08:02 Type: P Check: 9943 ..

1 MDVRIAVVG AGVGLSTAV CISKWPGS ITVISDKFTP EITSDVAAGM
 51 LIPTVPDTP IQKQKWEKE TFDHLEAIYN SAEADAGVI LVSGMQLFOS
 101 IPTPEVPPYA DVLGFRKMT KDELKFPQH VEGHATTLK CEGPAVLPWL
 151 QKRVKNGSL ILTRIDELW ELHPSDIYV NCSGLGSQL AGSKTFPVR
 201 GOYLKQAPW VKHTRDSSG LTIYVGVSN VTLGSTRQK DWNISPDARI
 251 SKELSRCCA LPSIRGAVD LREKVLGRLT RPSVLEKEL LAODSRRLPV
 301 VHHYGGSGG IAHMGTALE ATRLVNECVO VLRTPAKSK L

11AA_SEQUENCE 1.0 STANDARD; PRT: 341 AA.
 ID OXDD_HUMAN 099489; Created
 AC Q99489; 099489; (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DT 15-DEC-1999 (Rel. 39, Last annotation update)
 DE D-ASPARTATE OXIDASE (EC 1.4.3.1) (DASOX) (DDO).
 GN DDO.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE: 97306065.
 RA SETOYAMA C., MIURA R.;
 RT "Structural and functional characterization of the human brain D-
 RT aspartate oxidase".
 RT J. Biochem. 121:798-803(1997).
 RL -1- FUNCTION: SELECTIVELY CATALYZES THE OXIDATIVE DEAMINATION OF D-
 CC ASPARTATE AND ITS N-METHYLATED DERIVATIVE, N-METHYL D-ASPARTATE.
 CC -1- CATALYTIC ACTIVITY: D-ASPARTATE + H(2)O + O(2) = OXALACETATE +
 CC NH(3) + H(2)O(2).
 CC -1- COFACTOR: FAD OR 6-HYDROXYFLAVIN ADENINE DINUCLEOTIDE.
 CC -1- SUBCELLULAR LOCATION: PEROXISOMAL.
 CC -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS, DDO-1 (SHOWN HERE) AND DDO-2;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- SIMILARITY: BELONGS TO THE DAMOX/DASOX FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@lsb-sib.ch.

CC EMBL: D89858; BA014031.1; -
 CC HSSP: P00371; IDAO.
 DR MIM: 124450; -
 DR PROSITE: PS00677; DAO: 1.
 DR PFM: PF01266; DAO: 1.
 KW Oxidoreductase: Flavo-protein; FAD; Peroxisome; Alternative splicing.
 FT NP_BIND 6 20
 FT ACT_SITE 223 223 BY SIMILARITY.
 FT ACT_SITE 302 302 BY SIMILARITY.
 FT SITE 338 341 MICROBODY TARGETING SIGNAL (POTENTIAL).
 FT VARSITE 95 153 MISSING (IN ISOFORM DDO-2).
 SQ SEQUENCE 341 AA; 37535 MW; 0948D6AB CRC32;
 OXND_HUMAN Length: 341 February 14, 2000 08:02 Type: P Check: 7167 ..

1 MOTARIANV AGVGLSTAV CISKIVRCS VTIIDKFTP DTSDVAAGM
 51 LIPHTYDTP IHTOKWRE TENHPIRAN SAEADAGVH LVSGWQIFGS
 101 TPTEVPEFA DVIGFERMT EAIKKEPQY VEGQFTILK CECPALPML
 151 EKRIKGGW TLTRIEDLW ELHPSDIVV NCSGASROL AGSKIPVR
 201 GOVLOVOAPV VEHTRDSG LTYIPGTSH VTLGTRGKQ DMNLSPAEN
 251 SREILSRCA LEPSLHACN IREKVGRLPY RGVRLQTEL LARDGRLPY
 301 VHHYHSGSG ISVHWGTALE AARLVSECVH ALRPIPKSN L

11AA_SEQUENCE 1.0
 ID P2Y6_HUMAN STANDARD: PRT; 328 AA.
 AC 015077; 015754;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE P2Y PURINOCEPTOR 6 (P2Y6).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA COMMUNI D., PARMENTIER M., BOEYNAENS J.M.;
 RL Submitted (MAY-1996) to the EMBL/Genbank/DBD databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE: 97432828.
 RX SOMERS G.R., HAMMET F., WOOLLIATT E., RICHARDS R.I., SOUTHEY M.C.,
 RA VENTER D.J.;
 RT "Chromosomal localization of the human P2y6 purinoceptor gene and
 RT phylogenetic analysis of the P2y purinoceptor family.";
 RL Genomics 44:127-130(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE-PLACENTA;
 RX MEDLINE: 97415792.
 RA MAIER R., GLATZ A., MOSBACHER J., BILBE G.;
 RT "Cloning of P2y6 cDNAs and identification of a pseudogene: comparison
 RT of P2y receptor subtype expression in bone and brain tissues.";
 RL Biochem. Biophys. Res. Commun. 237:297-302(1997).
 RN [4]
 RP ERATUM.
 RX MEDLINE: 98069816.
 RA MAIER R., GLATZ A., MOSBACHER J., BILBE G.;
 RL Biochem. Biophys. Res. Commun. 240:298-302(1997).
 CC -1- FUNCTION: RECEPTOR FOR EXTRACELLULAR UDP > UTP > ATP. THE ACTIVITY
 CC OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE A
 CC PHOSPHATIDYLINOSITOL-CALCIUM SECOND MESSENGER SYSTEM.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>
 CC or send an email to license@lsb-sib.ch).

CC EMBL: X97058; CA65570.1; -
 CC EMBL: U52464; AAB03572.1; -
 CC EMBL: AF007892; AAB0713.1; -
 CC EMBL: AF007891; AAB0712.1; -
 CC HSSP: P34996; IDDD.
 DR MIM: 602451; -
 DR PROSITE: PS00237; G-PROTEIN-RECEPTOR; FALSE-NEG.
 DR PFM: PF00001; 7tm_1; 1.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein.
 FT DOMAIN 1 27
 FT TRANSSEM 28 48
 FT DOMAIN 49 62
 FT TRANSSEM 63 83
 FT DOMAIN 84 101
 FT TRANSSEM 102 122
 FT DOMAIN 123 144
 FT TRANSSEM 145 165
 FT DOMAIN 166 194
 FT TRANSSEM 195 215
 FT DOMAIN 216 236
 FT TRANSSEM 237 257
 FT DOMAIN 258 280
 FT TRANSSEM 281 303
 FT DOMAIN 304 328
 FT DIUPTID 99 177
 FT CARBOHYD 5 5
 FT CONFLICT 2 3
 SQ SEQUENCE 328 AA; 36429 MW; 29C125CE CRC32;
 P2Y6_HUMAN Length: 328 February 14, 2000 08:02 Type: P Check: 7545 ..

1 MEMDNGTGA LGLEPTTCY RENEKOLLP PYSAVLANG LPLNCVITO
 51 ICTSRALTR TAVYTINLAL ADLVACSLP LLIYNAQSD HWPEDDFCR
 101 LVRFLEFVNL HGSILFLTCI SFORYLGICH PLAPWKRKG RRAAMLVCA
 151 VWLAATTOCL PTAIFNATGI QNRRTVCYDL SPALATHW PYGALTYIG
 201 FLPPAALLA CYCLACRLC RODGPAEVA QERGRKARV AVVAAPAI
 251 SLPFHITKT AYLAVRSTPG VPCVLEAFA AAYKGRPFA SANVYLDPL
 301 FYTQKFRFR RPHLLQKLT AKWROGR

11AA_SEQUENCE 1.0
 ID P2Y6_HUMAN STANDARD: PRT; 166 AA.
 AC P31078;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE P2Y PURINOCEPTOR 6 (P2Y6).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN-MT1131;
 RX MEDLINE: 92356828.
 RA TOKITO M.K., DALDAL F.;
 RT "P2y6, located upstream of the fccR3 operon encoding the cytochrome
 RT bcl complex, is homologous to bacterial response regulators and
 RT necessary for photosynthetic and respiratory growth of Rhodospirillum rubrum
 RT capsulatus.";

RL MOL. Microbiol. 6:1645-1654(1992).
CC -1- FUNCTION: NECESSARY FOR PHOTOSYNTHETIC AND RESPIRATORY GROWTH.
CC -1- SIMILARITY: BELONGS TO THE MARR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 212113; CAN78097.1; -.
DR PIR: S22631; S22631.
DR PROSITE: PS01117; HTH_MARR_FAMILY; 1.
DR PFAM: PF01047; Marr; 1.
KW Transcription regulation; DNA-binding.
SO SEQUENCE 166 AA; 18807 MW; 99E3E86E CRC32;
PETP_RHOCA Length: 166 February 14, 2000 08:02 Type: P Check: 1875 ..
1 MADTGAPGGE TLFLTDEQL RKGIEAMFPA YRGFTADPDR ILDOHDYGRA
51 HRRATHTNR EPGLTVTLLI SYLGVTKOSI NRELTLLID GLVESRYGRR
101 DRRERHLHT EKGAVLEREL SEAGRVMPA AYRAGPOAV AGFROVLEAM
151 MDPARRHYQ MLKDAE
11AA SEQUENCE 1.0 STANDARD; PRT; 441 AA.
ID PEX3_YEAST
AC P28795.
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PEROXISOMAL MEMBRANE PROTEIN PAX3 (PEROXIN-3).
GN PEX3 OR PAX3 OR IDK329C OR D9798.15.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
NC [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91373453.
RA HOEHELD J., VEENHUIS M., KUNAU W.-H.;
RT "PAX3, a Saccharomyces cerevisiae gene encoding a peroxisomal
RT integral membrane protein essential for peroxisome biogenesis.";
RT J. Cell Biol. 114:1167-1178(1991).
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN: S288C / AB972;
RA JOHNSON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DU Z.,
RA FAYELLO A., HULTON L., GATUNG S., GRECO T., KIRSTEN J.,
RA KUCERA D., HALLSWORTH K., HARKINS J., HILLER L., JIR M.,
RA JOHNSON T., JOHNSON L., LANGSTON T., LATREILLE P., LE T.,
RA MARIIS E., MENEZES S., MILLER N., NHAN M., PAULEY A., PELUSO D.,
RA RIKKEN L., RILES L., TAICH A., TREVASKIS E., VIGNATI D.,
RA WITCOX L., WOLDMAN P., VAUDIN M., WILSON R., WATERSTON R.;
RL Submitted (Jul-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN PEROXISOME BIOSYNTHESIS. MAY FUNCTION
CC AS A RECEPTOR PROTEIN. LACK OF THIS PROTEIN CAUSES THE
CC PEROXISOMAL-DEFICIENT PHENOTYPE AND MISLOCALIZATION IN THE
CC CYTOSOL OF PEROXISOMAL MATRIX PROTEINS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PEROXISOMAL.
CC -1- INDUCTION: BY OLEIC ACID.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

CC EMBL: X58407; CAN41309.1; -.
DR EMBL: U32517; AAB64764.1; -.
DR PIR: A40550; A40550.
DR SGD: L0001339; PAX3.
KW Transmembrane; Peroxisome.
FT DOMAIN 1 17 MATRIX (POTENTIAL).
FT TRANSMEM 18 39 POTENTIAL.
FT DOMAIN 40 441 CYTOPLASMIC (POTENTIAL).
SO SEQUENCE 441 AA; 50675 MW; 96C48B96 CRC32;
PEX3_YEAST Length: 441 February 14, 2000 08:02 Type: P Check: 166 ..
1 MAPNQRSRL LQRRGKVL I SLTGIALFT TGSVVFFVK RMLYKQQLRI
51 TEQHFKEQI KRFEQTOED SLTIYELLP VYRWYLNEND LNDISYTOTL
101 KQOKNOLTPA KSSSRESSP LKSKAPLME LELKSLIV TVYTVSSLI
151 LTRIQNLTL TRNEYLDSAI KLTMQENCN KLQNFYMW TSWSDPEDK
201 ADDAWYMAAK KSKKEQGEV INEAFSLIS WMLNKGLS YNEITNQIE
251 IEFDGHPED TLTFEFSR LTNIFRNTS QIFQNNNNL TSILPKDSS
301 GQEFLLQTL DADALTSFHS NTLVFNQLVN ELTQCESTA TSIVLESILN
351 ESFHFIMNKV GIKTIANKRP GQEDQOQYOM AVFAMSKDC COEMLQTTAG
401 SSSGGSVNEY LATLDVQPL DDLASVYSN FGVSSEFSK P
11AA SEQUENCE 1.0 STANDARD; PRT; 246 AA.
ID PHB3_ALCOE
AC P14697.
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE ACETOACETYL-COA REDUCTASE (Ec 1.1.1.56).
GN PHB3.
OS Alkaligenes eutrophus.
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC Ralstonia.
NC [1]
RP SEQUENCE FROM N.A.
RX STRAIN: H16.
RX MEDLINE: 89359356.
RA PEOPLES O.P., SINSKEY A.J.;
RT "Poly-beta-hydroxybutyrate biosynthesis in Alkaligenes eutrophus H16.
RT Characterization of the genes encoding beta-ketothiolase and
RT acetoacetyl-CoA reductase.";
RT J. Biol. Chem. 264:15293-15297(1989).
RL -1- CATALYTIC ACTIVITY: (R)-3-HYDROXYACYL-COA + NADP(+) -> 3-OXOACYL-
CC COA + NADPH
CC -1- PATHWAY: SECOND STEP IN POLY-BETA-HYDROXYBUTYRATE BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC FAMILY (SDR).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J04987; AAA21973.1; -.
DR PIR: B34340; RDALAE.
DR HSSP: P19992; 2HSD.
DR PROSITE: PS00061; ADH_SHORT; 1.
DR PFAM: PF00106; adh_short_C2; 1.
DR PFAM: PF00678; adh_short_C2; 1.
KW Oxidoreductase; NADP; PHB biosynthesis.

FT NP_BIND 8 32 NADP (BY SIMILARITY).
 DR FRAM; PF00003; ABC_tran; 1.
 FT ACT_SITE 153 153 BY SIMILARITY.
 DR FRAM; PF00003; ABC_tran; 1.
 SQ SEQUENCE 246 AA; 26370 MW; B1A08AA3 CRC32;
 PHBB_ALCEU Length: 246 February 14, 2000 08:02 Type: P Check: 2829 ..
 1 MGRIVAVTG GMGGIGTAIC QRLAKDFRY VAGCGPNSPR REKMLEQOKA
 51 LGDFIASG NVADWDSTKT APDKASEVG EVDVLINAG ITRDVFRRKM
 101 TRADMVAVID TNLSTLFNVT KQYIDGMADR GWRIVNISS VNGKQGFQ
 151 TNSSTAKGL HGFTMALAE VAKRGVTVN VSPGIATDM VKAIRQVLD
 201 KIVATIPYKR LGPEELAST CAVLSESEG FSTGADPSLN GGLHMG
 11AA_SEQUENCE 1.0 STANDARD; PRT: 252 AA.
 ID PHNK_ECOLI
 AC 116678;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE PHOSPHONATES TRANSPORT ATP-BINDING PROTEIN PHNK.
 GN PHNK.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE; 91193228.
 RA MAKINO K., KIM S.K., SHINAGAWA H., AMEMURA M., NAKATA A.;
 RT "Molecular analysis of the cryptic and functional *phn* operons for
 RT phosphate use in *Escherichia coli* K-12."
 RL J. Bacteriol. 173:2665-2672(1991).
 RN 12
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE; 95334362.
 RA BURLAND V.D., PLUNKERT G. III, SOPHA H.J., DANIELS D.L.,
 RT "Analysis of the *Escherichia coli* genome VI: DNA sequence of the
 RT region from 92.8 through 100 minutes."
 RL Nucleic Acids Res. 23:2105-2119(1995).
 RN 13
 RP SEQUENCE FROM N.A.
 RC STRAIN-B;
 RX MEDLINE; 90170953.
 RA CHEN C.-M., YE Q.-Z., ZHU Z., WANNER B.L., WALSH C.T.;
 RT "Molecular biology of carbon-phosphorus bond cleavage. Cloning and
 RT sequencing of the *phn* (*psid*) genes involved in alkyldiphosphate
 RT uptake and C-P lyase activity in *Escherichia coli* B."
 RL J. Biol. Chem. 265:4461-4471(1990).
 CC -1- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
 CC FOR ALKYLPHOSPHONATES; PROBABLY RESPONSIBLE FOR ENERGY COUPLING
 CC TO THE TRANSPORT SYSTEM.
 CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
 CC (ABC TRANSPORTERS).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; D90227; AAA14271.1; -
 DR EMBL; U16003; AAA86986.1; -
 DR EMBL; AB000482; AAC77058.1; -
 DR EMBL; J05260; AAA24349.1; -
 DR PIR; C35719; C35719.
 DR ECOGENE; EG10720; PHNK.

DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
 DR FRAM; PF00003; ABC_tran; 1.
 FT NP_BIND 38 45 ATP (BY SIMILARITY).
 FT CONFLICT 47 45 L -> O (IN REF. 2).
 SQ SEQUENCE 252 AA; 27831 MW; CB559E63 CRC32;
 PHNK_ECOLI Length: 252 February 14, 2000 08:02 Type: P Check: 804 ..
 1 MNOPLSYNN LTHLYAPKRG FSDVSFDLMP GEVLGIAGES GSGKTTLLKS
 51 ISARLTPQOG EIHYENRSLY AMSEADRRRL LRTMGVYHO HPLDGLRRQV
 101 SAGGNIGERL MATGARHYGD IRATAKWLE EVEIPNRID DLPTFSGGM
 151 QORLOIANL VTHPKLVFMD EPTGLDVSV QARLLDLRG LVEELNAV
 201 IYTHDLGYAR LMDRLLYVK QGVVESGLT DRYLDPRHHR YQLLVSSVL
 251 ON
 11AA_SEQUENCE 1.0 STANDARD; PRT: 263 AA.
 ID P1A_ORISA
 AC 040708;
 DT 01-NOV-1987 (Rel. 35, Created)
 DT 01-NOV-1987 (Rel. 35, Last sequence update)
 DT 01-NOV-1987 (Rel. 35, Last annotation update)
 DE PIR/A PROTEIN.
 DE Oryza sativa (Rice).
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
 OC Poaceae; Oryza.
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. INDICA-IR36;
 RA REIMANN C., MAUCH F., DUDLER R., HOFMANN C.;
 RT "Characterization of a rice gene induced by *Pseudomonas syringae* pv.
 RT syringae: requirement for the bacterial *lema* gene function."
 RL Physiol. Mol. Plant Pathol. 46:71-81(1995).
 CC -1- SIMILARITY: STRONG, TO (5)-ACETONE-CYANOHYDRIN LYASE (EC
 CC 4.2.1.39).
 CC -1- SIMILARITY: DISTANT RELATIONSHIP WITH TYPE-B
 CC CARBOXYL ESTERASE/LIPASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; Z34271; CA84025.1; -
 DR HSP; P52704; IYAS.
 DR PFAM; PF00561; abhydrolase_1.
 FT ACT_SITE 82 82 BY SIMILARITY.
 FT ACT_SITE 213 213 BY SIMILARITY.
 FT ACT_SITE 241 241 BY SIMILARITY.
 SQ SEQUENCE 263 AA; 28663 MW; 0B95B213 CRC32;
 P1A_ORISA Length: 263 February 14, 2000 08:02 Type: P Check: 1305 ..
 1 MEDGGHNEF VHGLGYGAMC WYRVAAALRA AGHRAMALDM AAAGAHPARA
 51 DEVGSLSEYS RPLDDAVAAA APGERLVYVG HSLGGLSLAL AMERFPDKVA
 101 AAVFLAACMP AAGKHMGITL EEFMRIRKPD FFMDSKTIIV NTNQEPRTAV
 151 LGPKRLAEK LYNSRPEDL TLATMLVPRG TNYIDDPIMK DETLLREGNY
 201 GSVKRYFLVA MDAASDEEM QRWITDLSRG VEVELAGAD HMAKCKPRE

251 LCDLLLR1AA KYD

11AA_SEQUENCE 1.0
ID PLAS_SYNY3 STANDARD; PRT; 126 AA.
AC P21697;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, last sequence update)
DT 15-JUL-1998 (Rel. 36, last annotation update)
DE PLASTOCYANIN PRECURSOR.
GN PTE OR SL0109.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91338701.
RA BRIGGS L.M., PECORARO V.L., MCINTOSH L.;
RT "Copper-induced expression, cloning, and regulatory studies of the
RT plastocyanin gene from the cyanobacterium Synechocystis sp. PCC
RT 6803.";
RL Plant Mol. Biol. 15:633-642(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 96127529.
RA KANEKO T., TANAKA A., SATO S., KOTANI H., SAZUKA T., MIYAJIMA N.,
RA SUGIURA M., TABATA S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
RN [3]
RP SEQUENCE OF 29-47.
RX MEDLINE: 97443974.
RA SAZUKA T., OHARA O.;
RT "Towards a proteome project of cyanobacterium Synechocystis sp.
RT strain PCC6803: linking 130 protein spots with their respective
RT genes.";
RL Electrophoresis 18:1252-1258(1997).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS).
RX MEDLINE: 98128796.
RA ROMERO A., DE LA CERDA B., VARELA P.F., NAVARRO J.A., HERVAS M.,
RA DE LA ROSA M.A.;
RT "The 2.15 A crystal structure of a triple mutant plastocyanin from
RT the cyanobacterium Synechocystis sp. PCC 6803.";
RL J. Mol. Biol. 275:327-336(1998).
CC -1- FUNCTION: PLASTOCYANIN PARTICIPATES IN ELECTRON TRANSFER BETWEEN
CC -1- P700 AND THE CYTOCHROME B₆/F COMPLEX IN PHOTOSYSTEM I.
CC -1- INDUCTION: BY COPPER.
CC -1- SIMILARITY: CONTAINS ONE PLASTOCYANIN-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X54105; CAA38038.1; -
DR EMBL: D64000; BAA10227.1; -
DR PIR: S13733; S13733.
DR PDB: 1PCS; 17-DEC-97.
DR PROSITE: PS00196; COPPER-BIND. 1.
DR PFAM: PF00127; copper-bind. 1.
KW Electron transport; Copper; Signal; 3D-structure.
FT SIGNAL 1 28
FT CHAIN 1 28
FT DOMAIN 29 126 PLASTOCYANIN-
FT METAL 67 67 PLASTOCYANIN-LIKE.
FT METAL 111 111 COPPER.
FT METAL 114 114 COPPER.
FT METAL 119 119 COPPER.
SQ SEQUENCE 126 AA; 13146 MW; 0358BFF1 CRC32;

PLAS_SYNY3 Length: 126 February 14, 2000 08:02 Type: P Check: 6628

1 MSKRLTTLA GLLVYSSEF LSVSPAAN AYKMSDSG ALVEPSTVT
51 IKAGEYKMY NKLSPHNIV FAQCVADT AALSHKGLA PAGESFTST
101 FTEPETYTY CEPHRGAMV GKVWE

11AA_SEQUENCE 1.0
ID PLEK_HUMAN STANDARD; PRT; 350 AA.
AC P08567;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, last sequence update)
DT 15-DEC-1999 (Rel. 39, last annotation update)
DE PLECKSTRIN (PLATELET P47 PROTEIN).
GN PLEK OR P47.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 88232910.
RA TYERS M., RACHUBINSKI R.A., MCCAW M.L., VARRICHIO A.M.,
RA SHOR R.G.L., HASLAM R.J., HARLEY C.B.;
RT "Molecular cloning and expression of the major protein kinase C
RT substrate of platelets.";
RL Nature 333:470-473(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 89359547.
RA TYERS M., HASLAM R.J., RACHUBINSKI R.A., HARLEY C.B.;
RT "Molecular analysis of pleckstrin: the major protein kinase C
RT substrate of platelets.";
RL J. Cell. Biochem. 40:133-145(1989).
RN [3]
RP STRUCTURE BY NMR OF 1-105.
RX MEDLINE: 94268557.
RA YOON H.S., HADJUK P.J., PETROS A.M., OLEJNICZAK E.T., MEADOWS R.P.,
RA FESIK S.W.;
RT "Solution structure of a pleckstrin-homology domain.";
RL Nature 369:672-675(1994).
CC -1- FUNCTION: MAJOR PROTEIN KINASE C SUBSTRATE OF PLATELETS, ITS
CC EXACT FUNCTION IS NOT KNOWN.
CC -1- SIMILARITY: CONTAINS 2 PH DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X07743; CAA30564.1; -
DR PIR: S00755; S00755.
DR PIR: A45763; A45762.
DR PDB: 1PSS; 03-JUN-95.
DR MTM: 173570; -
DR PROSITE: PS0003; PH DOMAIN: 2.
DR PFAM: PF00169; PH. 2.
DR PFAM: PF00610; DEP. 1.
KW Phosphorylation; Repeat; 3D-structure.
FT DOMAIN 107 120 CONTAINS 3 POTENTIAL PHOSPHORYLATION
FT SITES.
FT CA_BIND 301 312 POTENTIAL.
FT DOMAIN 4 101 PH.
FT DOMAIN 244 347 PH.
FT VARIANT 92 92 W->R.
SQ SEQUENCE 350 AA; 40082 MW; 04F9FC1F CRC32;

PLEK_HUMAN Length: 350 February 14, 2000 08:02 Type: P Check: 1536

1 MEKRRREG LYKKSVENT WKPMVYLE DGEFFKKS DNSPKMPL
 51 KSLTSPQ DFKRMVEK ITTKQDHF FQALBERD AMVRLINKAI
 101 KCEGQKRA KSTRRSIRL PETIDGALY LSMKTEGI KEINLEKDK
 151 IFNHCTGNC VIDMLVNSQS VRRROECIMI ASSLINEGL QPAGDMKSA
 201 VGTAEENPL DNPATFYFP DSGFCEBNS SDDVILKEE FRYVITKQC
 251 LKQGHRRN WKVKFIRE DPAYLYHPD AGAEDPLCAI HLRCVTVS
 301 ESNNSGRKE EENLEIITA DEVHYELOA TPKEITEMIK AIOMASFTGK

11AA_SEQUENCE 1.0
 ID PLYA_MYCPO STANDARD; PRT; 299 AA.
 AC Q12639;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DE 15-DEC-1998 (Rel. 37, Last annotation update)
 GN PECA.
 OS Mycosphaerella pinodes (Pea foot rot fungus).
 OC Eukaryota; Fungi; Ascomycota; Euascomycetes; Loculoascomycetes;
 CC Dothideales; Mycosphaerellaceae; Mycosphaerella.
 RN (1)
 RP STRAIN-DSM 62763 / BERKELEY BLOXOM;
 RA HEIM P.;
 CC Submitted (MAY-1995) to the EMBL/Genbank/DBJ databases.
 CC 1. CATALYTIC ACTIVITY: ELIMINATIVE CLEAVAGE OF PECTIN TO GIVE
 CC OLIGOSACCHARIDES WITH TERMINAL 4-DEOXY-6-METHYL-ALPHA-D-GALACTO-
 CC 4-EPURONOSYL GROUPS.
 CC 1. SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.
 CC 1. SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

EMBL: X87580; CAA60884.1;
 DR PFM: PF00544; pec_lyase; 1.
 KW Lyase; signal.
 FT SIGNAL 1 18 POTENTIAL.
 FT CHAIN 19 299 PECTIN LYASE.
 SQ SEQUENCE 299 AA; 33072 MW; 121CF6C4 CRC32;

PLYA_MYCPO Length: 299 February 14, 2000 08:02 Type: P Check: 8743

1 MKSTPVSUG LTAITALAP TSPMIDYKRD VYKRAASLSD VAIERPLRC
 51 RPSASLEPPS RYTRLEPLFS LVLSRALATT SRAATSPS DRTARSLRT
 101 SPSPSRVSRM LSFVTPSKRR SVAMLSPSR RIELVTPKEI TTYLHRLQSE
 151 NWVWDLILS SDRDHDKDY DGLDITHAA DEVYTNFL HDHWKASLIG
 201 HSDSNGAEDK GHLVTVANN YLKNLSRGR PSASACATST TTTTRTCRMV
 251 STPARVRSCL FTTTSLAPR RLFTRLMDM LSLATSVW ARTLPRVL

11AA_SEQUENCE 1.0
 ID PME_ASPAC STANDARD; PRT; 331 AA.
 AC Q12535;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-1999 (Rel. 38, Last annotation update)
 GN PECTINESTERASE PRECURSOR (EC 3.1.1.11) (PECTIN METHYLESTERASE) (PE).
 DE

GN PME1.
 OS Aspergillus aculeatus.
 OC Eukaryota; Fungi; Ascomycota; Euascomycetes; Plecomycetes;
 CC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-KSM 510;
 RX MEDLINE: 97079238.
 RA CHRISTGAT S., KOFOED L.V., HALKIER T., ANDERSEN L.N., HOCKAUF M.,
 RA DORREICH K., DAUBORG H., KAUPIENEN S.;
 RT "Pectin methyl esterase from Aspergillus aculeatus: expression
 RT cloning in yeast and characterization of the recombinant enzyme."
 RL Biochem. J. 319:705-712(1996).
 CC 1. FUNCTION: INVOLVED IN MACERATION AND SOFT-ROT OF PLANT TISSUE.
 CC 1. CATALYTIC ACTIVITY: PECTIN N(2)O - N-METHANOL + PECTATE.
 CC 1. SIMILARITY: BELONGS TO THE PECTINESTERASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

EMBL: U49378; AAB42153.1;
 DR PROSITE: PS00800; PECTINESTERASE_1; 1.
 DR PROSITE: PS00503; PECTINESTERASE_2; 1.
 DR PFM: PF01095; Pectinesterase; 1.
 KW Hydrolase; Serine esterase; Cell wall; Signal.
 FT SIGNAL 1 17 BY SIMILARITY.
 FT CHAIN 18 331 PECTINESTERASE.
 SQ SEQUENCE 331 AA; 35681 MW; C567C2B3 CRC32;

PME_ASPAC Length: 331 February 14, 2000 08:02 Type: P Check: 2447

1 MKSVLAAL FNVSAASR TTPSGAIV AKSGDYTTI GAIDALSTS
 51 TTDITITIE EGYDEYVL PAMTGKVIIT GQENTDSYA DMLVITTHAI
 101 SYEDAGESD LTRATFNKAV GSOVYNLINA NTCQACHQA LALSAMADQ
 151 GYGCNFTGY QDTLLAOTGN QLYINSYIEG AVDFIFQHA RAMFQNVDIR
 201 VVEGPTASAI TANGRSSETD TSYVYINKST VAAKSGDVA EGYLYGRFW
 251 SEYAVVFOQ TSMTNVINSI GNTWSTSTP NTEVTFGEY ANTGAGEGT
 301 RASFAEKIDA KLITDILGS DYTSMVDTSY F

11AA_SEQUENCE 1.0
 ID PME_ASPAC STANDARD; PRT; 331 AA.
 AC P17872;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 GN PECTINESTERASE PRECURSOR (EC 3.1.1.11) (PECTIN METHYLESTERASE) (PE).
 DE

OS Aspergillus tubingensis.
 OC Eukaryota; Fungi; Ascomycota; Euascomycetes; Plecomycetes;
 CC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 RN (1)
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN-RH 5344;
 RX MEDLINE: 90332436.
 RA KHANH N.Q., ALBRECHT H., RUTKOWSKI E., LOEFFLER F., GOTTSCALK M.,
 RA JANY K.-D.;
 RT "Nucleotide and derived amino acid sequence of a pectinesterase cDNA
 RT isolated from Aspergillus niger strain RH 5344."
 RL Nucleic Acids Res. 18:4262-4262(1990).
 GN (2)
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 92039066.

RA KHANH N.O., RUTKOWSKI E., LEIDINGER K., ALBRECHT H., GOTSCHALK M.;
 RT "Characterization and expression of a genomic pectin methyl esterase-
 RL encoding gene in *Aspergillus niger*.";
 CC Gene 106:71-77(1991).
 CC -1- FUNCTION: INVOLVED IN MACERATION AND SOFT-ROTTING OF PLANT TISSUE.
 CC -1- CATALYTIC ACTIVITY: PECTIN + N H₂O = N METHANOL + PECTATE.
 CC -1- SIMILARITY: BELONGS TO THE PECTINESTERASE FAMILY.
 CC -1- CAUTION: STRAIN RH 5344 WAS PREVIOUSLY SAID TO BE FROM A NIGER.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: X52902; CAA37084.1; -
 DR EMBL: X54145; CAA38084.1; -
 DR PIR: S10487; S10487.
 DR PIR: J70589; J70589.
 DR PROSITE: PS00800; PECTINESTERASE_1; 1.
 DR PROSITE: PS00503; PECTINESTERASE_2; 1.
 DR PFAM: PF01095; Pectinesterase; 1.
 KW Hydrolyase; Serine esterase; Cell wall; Signal.
 FT SIGNAL 1 17
 FT CHAIN 18 331 PECTINESTERASE.
 SQ SEQUENCE 331 AA; 35715 MW; 80A4A5F0 CRC32;
 PME_ASPU Length: 331 February 14, 2000 08:02 Type: P Check: 609 ..

1 MKRSILASVL FAATATAASR MNPASGAIYV AKSGDDYTI SAADVALLST
 51 STETQIFIE EGSYDEQYI PALSGLIYV GQEDITTYT SNLVNTHAI
 101 ALADVNDDE TATLRNNAEG SAIYNININ TCGACRHAL AVASAYSEGG
 151 YVACQFTGQ DTLAETGQY VVAGITIEGA VDFIFGQAR AMHEDIRY
 201 LEGPSSASIT ANGRSESDS SYIVHKSTV AADGNDVSS GIYILGRPS
 251 QYAVCFQKT SMVDVNHLG WTEWSTSTPN TENVFEVYG NGTGAEGPR
 301 ANFSELTPE ITISMLGSD MEDWDTSYI N
 11AA_SEQUENCE 1.0
 ID PSPB_BOVIN STANDARD; PRT; 79 AA.
 AC P15781;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-NOV-1990 (Rel. 15, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE PULMONARY SURFACTANT-ASSOCIATED PROTEIN B (SP-B) (6 KD PROTEIN)
 DE (PULMONARY SURFACTANT-ASSOCIATED PROTEOLIPID SPL(PHE)).
 GN SPFB OR SPFB3.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
 OC Bovinae; Bos.
 RN [1]
 RP MEDLINE: 88077030.
 RA OLAFSON R.W., RINK U., KIELAND S., YU S.-H., CHUNG J.,
 RA HARDING P.G.R., POSSWAYER E.;
 RT "Protein sequence analysis studies on the low molecular weight
 RT hydrophobic proteins associated with bovine pulmonary surfactant.";
 RL Biochem. Biophys. Res. Commun. 148:1406-1411(1987).
 RN [2]
 RP MEDLINE: 88025156.
 RA YU S.-H., CHUNG W., OLAFSON R.W., HARDING P.G.R., POSSWAYER E.;
 RT "Characterization of the small hydrophobic proteins associated with
 RT pulmonary surfactant.";
 RL Biochim. Biophys. Acta 921:437-448(1987).

CC -1- FUNCTION: PULMONARY SURFACTANT ASSOCIATED PROTEINS PROMOTE
 CC ALVEOLAR STABILITY BY LOWERING THE SURFACE TENSION AT THE AIR-
 CC LIQUID INTERFACE IN THE PERIPHERAL AIR SPACES. SP-B INCREASES
 CC THE COLLAPSE PRESSURE OF PALMITIC ACID TO NEARLY 70 MILLINERTONS
 CC PER METER.
 CC -1- SUBUNIT: HOMODIMER, DISULFIDE-LINKED.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -1- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%
 CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
 CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
 CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).
 DR PIR: A29667; A29667.
 DR PIR: S02317; S02317.
 KW Surface film; Gaseous exchange.
 SQ SEQUENCE 79 AA; 8660 MW; 2B73807E CRC32;
 PSPB_BOVIN Length: 79 February 14, 2000 08:02 Type: P Check: 7861 ..

1 FPIPIYCW LRLIKKIOA VIPKGVLAAT VAQVHVPL LVGSIIOOLV
 51 IEYSVILXTD TLGLRPLNV CGLRLRCSG
 11AA_SEQUENCE 1.0
 ID PTB_BACSU STANDARD; PRT; 299 AA.
 AC P54530;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE PROBABLE PHOSPHATE BUTYRYLTRANSFERASE (EC 2.3.1.19)
 DE (PHOSPHOTRANSBUTYRYLASE).
 GN Y01S
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC KOBAYASHI Y., MIZONO M., MASUDA S., TAKEMARU K., HOSONO S.,
 RA SATO T., TAKEUCHI M.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CATALYZES THE CONVERSION OF BUTYRYL-COA THROUGH BUTYRYL
 CC PHOSPHATE TO BUTYRATE.
 CC -1- CATALYTIC ACTIVITY: BUTANOYL-COA + ORTHOPHOSPHATE = COA +
 CC BUTANOYLPHOSPHATE.
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHATE ACETYLTRANSFERASE AND
 CC BUTYRYLTRANSFERASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: D84432; BAA12594.1; -
 DR EMBL: Z99116; CAB14340.1; -
 DR SUBTILIST; BG11722; Y01S.
 DR PFAM: PF01515; PTA_PTB; 1.
 KW Hypothetical protein; Transferase; Acyltransferase.
 SQ SEQUENCE 299 AA; 31772 MW; EF3F522D CRC32;
 PTB_BACSU Length: 299 February 14, 2000 08:02 Type: P Check: 158 ..

1 MKLKDIGKA SIHKKTIAY AHAEDEVIR AVKLAHEHLS APRLLTGDSK
 51 KLELTSWQ GHVEIYHAN TPESAKLAY RAVHKRTADV LMGVPTSV
 101 LKAVLNROE GLRSASYLSH VAVEDIPDD RLMFVTSAM NNPSELELR
 151 QIQQNVVHA HAVGNMKA ALAAVEYVA PKMEATVAA ALQMKRRRO
 201 IKCIYDGL ALDNNVSGIA AAKKISGV AGNDILLV TTAGNITLK

251 SLIFAKASY AAVITGAKAP IALTSRADSA ENKLYSIALA ICASEETH

11AA_SEQUENCE 1.0 STANDARD: PRT: 376 AA.
 ID PT16_HUMAN
 AC P35237;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, last sequence update)
 DT 01-OCT-1996 (Rel. 34, last annotation update)
 DE PLACENTAL THROMBIN INHIBITOR (CYTOSOLASMIC ANTIPTROTEINASE) (CAP)
 DE (PROTEASE INHIBITOR 6).
 GN P16 OR P11.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-PLACENTA;
 RX MEDLINE: 94022386.
 RA COUGHLIN P., SUN J., CERRUTI L., SALEM H.H., BIRD P.;
 RT "Cloning and molecular characterization of a human intracellular
 serine proteinase inhibitor.";
 RT Proc. Natl. Acad. Sci. U.S.A. 90:9417-9421(1993).
 RL [2]
 RN SEQUENCE FROM N.A., AND SEQUENCE OF 67-73 AND 144-149.
 RP TISSUE-PLACENTA;
 RX MEDLINE: 94183847.
 RA MORENSTERN K.A., SPRECHER C.A., HOLTH L., FOSTER D., GRANT F.J.,
 CHING A., KISIEL W.;
 RT "Complementary DNA cloning and kinetic characterization of a novel
 intracellular serine proteinase inhibitor: mechanism of action with
 trypsin and factor xa as model proteinases.";
 RT Biochemistry 33:3432-3441(1994)
 CC -1- SUBCELLULAR LOCATION: CYTOSOLASMIC
 CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS IN SKELETAL MUSCLE, ALSO FOUND
 CC IN PLACENTA, CARDIAC MUSCLE, LUNG, LIVER, KIDNEY AND PANCREAS.
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: 222658; CA80373.1; -
 DR EMBL: 569272; AA830320.1; -
 DR PIR: S35750; S35750.
 DR PIR: A48681; A48681.
 DR HSSP: P05619; 1HLE.
 DR MIM: 173321; -
 DR PROSITE: PS00284; SERPIN; 1.
 DR PFAM: PF00079; serpin; 1.
 KW Serpin. Serine protease inhibitor.
 FT ACT_SITE 341 342 REACTIVE BOND.
 FT ACT_SITE 175 175 G -> E (IN REF. 2).
 FT CONFLICT 362 362 R -> S (IN REF. 2).
 SO SEQUENCE 376 AA; 42587 MW; 2D806418 CRC32.

PT16_HUMAN Length: 376 February 14, 2000 08:02 Type: P Check: 300

1 MDLCEANGT FALNLTITG KDNKRVFFS PKMSCALAM VYMGAGNNA

51 AQAQALISLN KSGGGDIHQ GFOSLLEVN KTGTQYLLRV ANLFGKSC

101 DFLSSFRDSC OKFYQAEKEE LDFISAVEKS RKHINTVAE KTGGKIAELL

151 SPQSDPLTR LVLYNAVYFR GNDGQFDKE NTEERLEKVS KNEKPYQAK

201 FKQSTFKKTY IGEITQILV LPYVGKELNA IIMDPDETD LRTVEKELY

251 EKFEVETRLD WDEDEVEVS LPRFKLEST DMESVLRNLG MTDAPFLGKA

301 DFGMSQIDL SLISKVXSF VEVEEGTEA AATPAIMMM RCARFVPRFC

11AA_SEQUENCE 1.0 STANDARD: PRT: 374 AA.
 ID PT18_HUMAN
 AC P50452;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, last sequence update)
 DT 01-NOV-1997 (Rel. 35, last annotation update)
 DE CYTOSOLASMIC ANTIPTROTEINASE 2 (CAP2) (PROTEASE INHIBITOR 8).
 GN P18
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-PLACENTA;
 RX MEDLINE: 96102039.
 RA SPRECHER C.A., MORENSTERN K.A., MATHEWS S., DAHLIN J.R.,
 SCHRAEDER S.K., FOSTER D.C., KISIEL W.;
 RT "Molecular cloning, expression, and partial characterization of two
 novel members of the ovalbumin family of serine proteinase
 inhibitors.";
 RT J. Biol. Chem. 270:29854-29861(1995).
 RL J. Biol. Chem. 270:29854-29861(1995).
 CC -1- SUBCELLULAR LOCATION: CYTOSOLASMIC.
 CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY. OV-SERPIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: I40377; AAC41939.1; -
 DR HSSP: P05619; 1HLE.
 DR MIM: 601697; -
 DR PROSITE: PS00284; SERPIN; 1.
 DR PFAM: PF00079; serpin; 1.
 KW Serpin. Serine protease inhibitor.
 FT ACT_SITE 339 340 REACTIVE BOND (BY SIMILARITY).
 SO SEQUENCE 374 AA; 42786 MW; 79AF74BC CRC32.

PT18_HUMAN Length: 374 February 14, 2000 08:02 Type: P Check: 4145

1 MDLCEANGT FALNLTITG EDNSRVFF SPMSISSALA VYMGAGNST

51 AQAQALISLN KYDGDHNGF QSLSEVNRT GTQYLLRTAN RLFEKTCDF

101 LPDFREYQK FYQAELEELS FAEDTECRK HINDVAEKT EGISSEVLDA

151 GTVDPILRLV LVNAVYFKK WNEQFDKRYT RGNLEFTNEE KKTVOEMKE

201 AKFKMGYADE VHTQVLELP VEELSMVIL LPDDNTDLAV VEKALTEKX

251 KAMNSEKLT KSKVQYFLPR LKLESYDLE PFIRRLGMD AFDEKADFS

301 GMSTERNVPL SKVAHKCFYE VNEEGTEAAA ATRVENSRC SRMEPRFCAD

351 HPFLFFIRRH KTNCLIFCGR FSSP

11AA_SEQUENCE 1.0 STANDARD: PRT: 248 AA.
 ID PUR7_METH
 AC O26272;
 DT 15-DEC-1999 (Rel. 39, Created)
 DT 15-DEC-1999 (Rel. 39, last sequence update)
 DT 15-DEC-1999 (Rel. 39, last annotation update)
 DE PHOSPHORIBOSYLTRANSFERINADOLE-SOCCINOCARBOXAMIDE SYNTHASE (EC 6.3.2.6)
 DE (SAICAR SYNTHETASE).
 GN PORC OR MTH10.

OS Methanobacterium thermoautotrophicum.
 OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
 OC Methanobacterium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DELTA H;
 RX MEDLINE: 98037514.
 RA SMITH D.R., DOUCETTE-STAMM L.A., DELOUGHERY C., LEE H.-M., DOBOIS J.,
 RA ALBREDE T., BASHIRZADEH R., BLAKELY D., COOK R., GILBERT K.,
 RA HARRISON D., HOANG L., KEAGLE P., LUMM W., POTNER B., QIU D.,
 RA SPADAPORA R., VICARE R., WANG Y., WIERZBOWSKI J., GIBSON R.,
 RA JIVANI N., CARUSO A., BUSH D., SAFER H., PATWELL D., PASHAKAR S.,
 RA MCDUGALL S., SHIMER G., GOYAL A., PIETROVSKI S., CHURCH G.M.,
 RA DANIELS C.J., MAO J.-I., RICE P., NOLLING J., REEVE J.N.,
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum
 RT deltaH: functional analysis and comparative genomics.";
 RL J. Bacteriol. 179:7135-7153(1997).
 CC -1- CATALYTIC ACTIVITY: ATP + 1-(5-PHOSPHORIBOSYL)-4-CARBOXY-5-
 CC AMINOIMIDAZOLE + L-ASPARTATE = ADP + ORTHOPHOSPHATE + 1-(5-
 CC PHOSPHORIBOSYL)-4-(N-SUCCINO-CARBOXYAMIDE)-5-AMINOIMIDAZOLE.
 CC -1- PATHWAY: SEVENTH STEP IN DE NOVO PURINE BIOSYNTHESIS.
 CC -1- SIMILARITY: BELONGS TO THE SAICAR SYNTHETASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-std.ch/announce/>
 CC or send an email to license@isb-std.ch).
 CC -----
 DR EMBL: AE000804; AAB84576.1;
 DR PROSITE: PS01057; SAICAR_SYNTHETASE_1; 1.
 DR PROSITE: PS01058; SAICAR_SYNTHETASE_2; 1.
 DR PFAM: PF01259; SAICAR_synct; 1.
 DR Purine biosynthesis; ligase.
 SQ SEQUENCE 248 AA; 28244 MW; 05FA27E0 CMC32;
 PUR7_METH Length: 248 February 14, 2000 08:02 Type: P Check: 5550
 1 MDVKIDGPLY SGRKADVLT DPDEIVAVRF RDTITADGE KDTLEMKY
 51 YNSVISAKIF EVLEAGVPT OYLELEPQC ILARKLEMP IEVITRNIAA
 101 GSIVRPFPT EGGEFVPLI QMDYKSDENG DPMINDIIL ALGIATRDEL
 151 EIRITLTHI NSVLRFILKS RGLILPDKL EFGDSSGRI RLGDVSPPT
 201 CRLMNMGE PLDKIFRRG EGVGVAVNR VARILDED IERNWEL
 11AA SEQUENCE 1.0 STANDARD; PRT; 310 AA.
 ID PYRB_ECOLI Q47555; Q47557
 AC P00479; Q47555; Q47557
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-FEB-1986 (Rel. 31, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ASPARTATE CARBAMOYLTRANSFERASE CATALYTIC CHAIN (EC 2.1.3.2) (ASPARTATE
 DE TRANS-CARBAMYLASE) (ATCASE).
 GN PYRB.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 84119419.
 RA SCHACHMAN H.K., PAUZA C.D., NAVRE M., KARELS M.J., WU L., YANG Y.R.;
 RT "Location of amino acid alterations in mutants of aspartate
 RT transcarbamoylase: structural aspects of interallelic
 RT complementation";
 RL Proc. Natl. Acad. Sci. U.S.A. 81:115-119(1984).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 83195078.
 RA WILD J.R.;
 RT "Nucleotide sequence of the structural gene (pyrB) that encodes the
 RT catalytic polypeptide of aspartate transcarbamoylase of Escherichia
 RT coli.";
 RL Proc. Natl. Acad. Sci. U.S.A. 80:2462-2466(1983).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE: 9534362.
 RA BURLAND V.D., PLUNKETT G. III, SOFIA H.J., DANIELS D.L.,
 RA BLATTNER F.R.,
 RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
 RT region from 92.8 through 100 minutes";
 RL Nucleic Acids Res. 23:2105-2119(1995).
 RN [4]
 RP REVISION TO 195.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE: 9742617.
 RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
 RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
 RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
 RA MAU B., SHAO Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [5]
 RP SEQUENCE.
 RX MEDLINE: 83195079.
 RA KONGSBERG N.H., HENDERSON L.;
 RT "Amino acid sequence of the catalytic subunit of aspartate
 RT transcarbamoylase from Escherichia coli.";
 RL Proc. Natl. Acad. Sci. U.S.A. 80:2467-2471(1983).
 RN [6]
 RP SEQUENCE OF 1-53 FROM N.A.
 RX MEDLINE: 89017155.
 RA ROLAND K.L., LIU C., TURNBOUGH C.L. JR.;
 RT Role of the ribosome in suppressing transcriptional termination at
 RT the pyrBI attenuator of Escherichia coli K-12.";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:7149-7153(1988).
 RN [7]
 RP SEQUENCE OF 1-45 FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE: 85289046.
 RA ROLAND K.L., POWELL F.E., TURNBOUGH C.L. JR.;
 RT "Role of translation and attenuation in the control of pyrBI operon
 RT expression in Escherichia coli K-12.";
 RL J. Bacteriol. 163:991-999(1985).
 RN [8]
 RP SEQUENCE OF 1-45 FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE: 83169660.
 RA TURNBOUGH C.L. JR., HICKS K.L., DONAHUE J.P.;
 RT "Attenuation control of pyrBI operon expression in Escherichia coli
 RT K-12.";
 RL Proc. Natl. Acad. Sci. U.S.A. 80:368-372(1983).
 RN [9]
 RP SEQUENCE OF 1-17 FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE: 91035438.
 RA DONAHUE J.P., TURNBOUGH C.L. JR.;
 RT Characterization of transcriptional initiation from promoters p1 and
 RT p2 of the pyrBI operon of Escherichia coli K12.";
 RL J. Biol. Chem. 265:19091-19099(1990).
 RN [10]
 RP SEQUENCE OF 1-11.
 RC STRAIN-K12 / W3110;
 RA PASQUALLI C., SANCHEZ J.-C., RAVIER F., GOLAZ O., HUGHES G.J.,
 RA FERTIGER S., PAQUET N., WILKINS M., APPEL R.D., BAIRICH A.,
 RA HOCHSTRASSER D.F.,
 RT Submitted (Sep-1994) to the SWISS-PROT data bank.
 RL [11]
 RP SEQUENCE OF 1-20.
 RC STRAIN-K12 / EMG2;

Mon Feb 14 08:07:43 2000

sp.cat

Page 102

FT HELIX 285 304
SO SEQUENCE 310 AA; 34296 MW; BDF9E927 CRC32;
PYRB_ECOLI Length: 310 February 14, 2000 08:02 Type: P Check: 3945 ..

1 ANPLYORHII SINDLSRDL NLYLATAKL KANPOPELAK HKVIASCFEE
51 ASTRTRLSFE TSMHRLGASV VGFSDSANTS LGKKGETLAD TIVISTRYD
101 AIVARHPOEG AARLATEFSG NVPLVINGDG SNQHPQTL DLFIOETOG
151 RLDELHYAVM GDLKYGRTV SLTQALAKFD GNFETIAD ALAPQYILD
201 MLDEKGIAMS LHSIEEVYA EVDILYMTV OKERLDPSEY ANYACQYLR
251 ASDLHNKAK MKVHLPLPRV DEITVDYKT PHAWTFQAG NGIFARQALL
301 ALVYNRLVYL

11AA_SEQUENCE 1.0 STANDARD; PRT; 319 AA.
ID PYRB_MYCTU
AC P71808;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE ASPARTATE CARBAMOYLTRANSFERASE (EC 2.1.3.2) (ASPARTATE
DE TRANSCARBAMYLASE) (ATCASE).
GN PYRB OR RV1380 OR MTCY02B12.14.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-H37RV;
RC MEDLINE: 98295987.
RA COLE S.T., BROSCHE R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA GORDON S.V., EIGLMEIER K., GAS S., BARRY C.E., III, TERKHA F.,
RA BADOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
RA DAVIES R., DEVLIN K., FELTWEIL T., GENTLES S., HAMLIN N., HOLROYD S.,
RA HORSLEY T., JAGELS K., KROGH A., MCLEAN J., MUIRE S., MURPHY L.,
RA OLIVER S., OSBORNE J., QUAIL M.A., RAINDEAN M.A., ROGERS J.,
RA RUTER S., SEGER K., SKELTON S., SQUARES S., SOKES R., STURSTON J.E.,
RA TAYLOR K., WHITEHEAD S., BARRELL B.G.,
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
CC -1- CATALYTIC ACTIVITY: CARBAMOYL-PHOSPHATE + ASPARTATE -
CC -1- ORTHOPHOSPHATE + N-CARBAMOYLASPARTATE.
CC -1- PATHWAY: SECOND STEP IN PYRIMIDINE BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE ATCASES/OTCASES FAMILY.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z81011; CAB02641.1; -
CC DR HSSP: P00479; 3AT1.
CC DR PROSITE: PS00097; CARBAMOYLTRANSFERASE; 1.
CC DR PFAM: PF00185; OTCase; 1.
CC KW Pyrimidine biosynthesis; Transferrase.
CC SO SEQUENCE 319 AA; 33818 MW; 38CF9A86 CRC32;

PYRB_MYCTU Length: 319 February 14, 2000 08:02 Type: P Check: 1547 ..

1 MTPHILTLAA DLSRDATAI LDDADRAQA LVGRDIKLP TLKGRIVYM
51 FYENSTRTV SPEVAGKMS ADVINVSAG SSVGKESIR DFLALLRAAG
101 ADALITHPA SGAAHLLAOM TGAHNDGPV INAGDGTHER PTOALLDALI

151 IRQRLGIEG RIYIVGDL HSRVANSVM LDTLGAEEV LVAPILLPV
201 GYTGMPATVS HPEDELPPAA DAVILMRVA ERMNGGFPS VREYSVRYGL
251 TERQAMLPQ HAVVHPGPM VRGMEITISV ADSOSAVIQ QVNGVQVRM
301 ALVFLVAVGA ODAKREGAA

11AA_SEQUENCE 1.0 STANDARD; PRT; 308 AA.
ID PYRB_PYRAB
AC P77918;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ASPARTATE CARBAMOYLTRANSFERASE (EC 2.1.3.2) (ASPARTATE
DE TRANSCARBAMYLASE) (ATCASE).
GN PYRB.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-GE5.
RC MEDLINE: 97352668.
RA PURCAREA C., HERVE G., LADJANI M.M., CUNIN R.,
RT "Aspartate transcarbamylase from the deep-sea hyperthermophilic
RT archaeon Pyrococcus abyssi: genetic organization, structure, and
RT expression in *Escherichia coli*.";
RL J. Bacteriol. 179:4143-4157(1997).
CC -1- CATALYTIC ACTIVITY: CARBAMOYL-PHOSPHATE + ASPARTATE -
CC -1- ORTHOPHOSPHATE + N-CARBAMOYLASPARTATE.
CC -1- PATHWAY: SECOND STEP IN PYRIMIDINE BIOSYNTHESIS.
CC -1- SUBUNIT: CONTAINS CATALYTIC AND REGULATORY CHAINS.
CC -1- SIMILARITY: BELONGS TO THE ATCASES/OTCASES FAMILY.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U61765; AAB62984.1; -
CC DR HSSP: P00479; 3AT1.
CC DR PROSITE: PS00097; CARBAMOYLTRANSFERASE; 1.
CC DR PFAM: PF00185; OTCase; 1.
CC KW Pyrimidine biosynthesis; Transferrase.
CC SO SEQUENCE 308 AA; 34901 MW; 7CEC7D77 CRC32;

PYRB_PYRAB Length: 308 February 14, 2000 08:02 Type: P Check: 9304 ..

1 MMKRRDVIS IRPFSKEDIE TVLNTAELE REIKERGOLE YAKKILIALI
51 FPEBTRRL SFESAMHRIQ GAVIGFAEAS TSSVKKGESL RDTIKTYEY
101 CDVYIHRPK EGARLADEV AVEVINAGD GSNQHPQTL LDLYTIKKEF
151 GRIDGLIKGL LSDLKYGRTV HSLAEALTFY DVELYLISPE LLNRPRIIVE
201 ELREKMKVY ETTLEDVIG KLDVLYTRI OKERPEDEDE YLKVGSYOV
251 NLKYLAKAD ELRIMHPLR VDEIHPEVDN TTHATYFROV FNGVPVMAI
301 LALVIGVI

11AA_SEQUENCE 1.0 STANDARD; PRT; 308 AA.
ID PYRB_PYRHO
AC O58451;
DT 15-DEC-1999 (Rel. 39, Created)
DT 15-DEC-1999 (Rel. 39, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE ASPARTATE CARBAMOYLTRANSFERASE (EC 2.1.3.2) (ASPARTATE

DE TRANSCARBAMYLASE) (ATCASE).

GN PYRB OR PH0720.

OS Pyrococcus horikoshii.

OC Archaea: Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-OT3.

RA MEDLINE: 96344137.

RA KAMARABAYATI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,

RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOTAKA A., NAGAI Y.,

RA SAKAI M., OGURA K., OTSUKA R., NAKAZAWA H., TAKAMITA M., OHFUKU Y.,

RA FUNAHASHI T., TANAKA T., KODOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,

RA AOKI K., NAKAMURA Y., ROBE T.F., HORIKOSHI K., MASUCHI Y., SHIZUTA H.,

RA KIKUCHI H.;

RT "Complete sequence and gene organization of the genome of a hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";

RL DNA Res. 5:55-76(1998).

CC -1- CATALYTIC ACTIVITY: CARBAMOYL-PHOSPHATE + ASPARTATE =

CC ORTHOPHOSPHATE + N-CARBAMOYL-ASPARTATE.

CC -1- PATHWAY: SECOND STEP IN PYRIMIDINE BIOSYNTHESIS.

CC -1- SUBUNIT: CONTAINS CATALYTIC AND REGULATORY CHAINS.

CC -1- SIMILARITY: BELONGS TO THE ATCASES/OTCASES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC EMBL: AP000003; BAA28611.1.

DR PROSITE: PS00097; CARBAMOYLTRANSFERASE; 1.

DR PFAM: PF00185; OTCase; 1. Transferase.

KW Pyrimidine biosynthesis; 1263350D CRC32;

SQ SEQUENCE 308 AA; 34862 MW; 1263350D CRC32;

PYRB_PYRHO Length: 308 February 14, 2000 08:02 Type: P Check: 9980 ..

1 MEMKGRDVIS IIRDFSKEDIE VVISTARLE KEMKRGOLE YAKKRIIATL

51 FEPSTRTRL SPESAMRLG GSVIGFVAS TSVKKGEST RDITKTEVOY

101 SDYIVIRHPK EGAARLAEV ADIPVINGD GSNQHPOTLL LDLYTIKKEF

151 GTIDGLKIGL LGDLKGRIV HSLAEALAFY DVELYLISPE LLRMPKHIVE

201 ELERGRMKIV ETTKLEEVIG ELVDLYYTRI QKERPDDEE YLKVGSYOV

251 NKLILEVND SURIMHPLPR VDEIHPEVDK TKHAIYFKOV FNGVPVBMAL

301 LALVLYGI

11AA_SEQUENCE 1.0 STANDARD; PRT; 310 AA.

ID PYRB_SALTY

AC P08420;

DT 01-AUG-1988 (Rel. 08, Created)

DT 01-FEB-1991 (Rel. 17, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE ASPARTATE CARBAMOYLTRANSFERASE CATALYTIC CHAIN (EC 2.1.3.2) (ASPARTATE TRANS-CARBAMYLASE) (ATCASE).

OS Salmoella typhimurium.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-LT2;

RA MEDLINE: 87246692.

RA MICHAELS G., KELIN R.A., NANGANG F.E.;

RT "Cloning, nucleotide sequence and expression of the pyrBI operon of *Salmoella typhimurium* LT2.";

RL Eur. J. Biochem. 166:55-61(1987).

CC -1- CATALYTIC ACTIVITY: CARBAMOYL-PHOSPHATE + ASPARTATE =

CC ORTHOPHOSPHATE + N-CARBAMOYL-ASPARTATE.

CC -1- PATHWAY: SECOND STEP IN PYRIMIDINE BIOSYNTHESIS.

CC -1- SUBUNIT: HETERODODECAMER (2C3:3R2) OF SIX CATALYTIC PYRB CHAINS ORGANIZED AS TWO TRIMERS (C3), AND SIX REGULATORY PYRI CHAINS ORGANIZED AS THREE DIMERS (R2).

CC -1- SIMILARITY: BELONGS TO THE ATCASES/OTCASES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC EMBL: X05641; CA29129.1.

DR PIR: S00049; OMEBAC.

DR HSSP: P00479; IRAG.

DR STYGENE: SG10324; PYRB.

DR PROSITE: PS00097; CARBAMOYLTRANSFERASE; 1.

DR PFAM: PF00185; OTCase; 1.

KW Pyrimidine biosynthesis; Transferase.

FT INIT MET 0 BY SIMILARITY.

SQ SEQUENCE 310 AA; 34307 MW; 01405B57 CRC32;

PYRB_SALTY Length: 310 February 14, 2000 08:02 Type: P Check: 2563 ..

1 ANPLVOKHII SINDLSRDL NLVLTAKL KANQPELLK HVIASCFEE

51 ASTRRLSFE TSMHRLGASV VGFSDSANTS LGRKGETLAD TISVISTYVD

101 AIVMHPQEG AARLATEFSG QVPVLNAGDG SNQHPOTLL DLFTIOETOG

151 RLDNHIIMV GDLYKGTIV FAKPRTIAKF SGNFYIAP DLAAMPQYLL

201 DMUDEGQAV SLHSISTEVA ADVLLIYTR YQKERDPSE YANVAQOPL

251 RPDNGAREN MKVHPLPRI DETTDVDTK PHANYFOQAG NGIFAAQALL

301 ALVINSELSL

11AA_SEQUENCE 1.0 STANDARD; PRT; 305 AA.

ID PYRB_SERMA

AC P19910;

DT 01-FEB-1991 (Rel. 17, Created)

DT 01-FEB-1991 (Rel. 17, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE ASPARTATE CARBAMOYLTRANSFERASE CATALYTIC CHAIN (EC 2.1.3.2) (ASPARTATE TRANS-CARBAMYLASE) (ATCASE).

GN Serratia marcescens.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Serratia.

RN [1]

RP SEQUENCE FROM N.A.

RA BECK D., KEDZIE K.M., WILD J.R.;

RT "Comparison of the aspartate transcarbamoylases from *Serratia marcescens* and *Escherichia coli*.";

RL J. Biol. Chem. 264:16629-16637(1989).

CC -1- CATALYTIC ACTIVITY: CARBAMOYL-PHOSPHATE + ASPARTATE =

CC ORTHOPHOSPHATE + N-CARBAMOYL-ASPARTATE.

CC -1- PATHWAY: SECOND STEP IN PYRIMIDINE BIOSYNTHESIS.

CC -1- SUBUNIT: HETERODODECAMER (2C3:3R2) OF SIX CATALYTIC PYRB CHAINS ORGANIZED AS TWO TRIMERS (C3), AND SIX REGULATORY PYRI CHAINS ORGANIZED AS THREE DIMERS (R2).

CC -1- SIMILARITY: BELONGS TO THE ATCASES/OTCASES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: J05033; AAA26564.1; -

DR PIR: B34396; OMSEC.

DR HSSP: P00479; 1RAG.

DR PROSITE: PS00097; CARBAMOYLTRANSFERASE; 1.

DR PFAM: PF00185; OTCace; 1.

KW Pyrimidine biosynthesis; Transferase.

FT INIT MET 0 BY SIMILARTY

SEQUENCE 305 AA; 33240 MW; 3E4D883A CRC32;

PIRB_SERNA Length: 305 February 14, 2000 08:02 Type: P Check: 4826 ..

1 ANPLHKHII SINDLRDL ELVLAPAGI KANPOPELHK HVIASCFEE

51 ASTRTLSFE TSMHRLGASV VGFADGNTS LKKKGETLAD TISVISTYVD

101 AIVMRHPDGG ARMASESGN VVPLNAGDN QHPITLIDL FTIQTQGR

151 SNLSIAMVGD LKGRVYHSL TQALAFEGN RFFIAPDAL AMPAYILKML

201 EEKGIYSSSH GSIEVYVEL DIYYMTRVOK ERLDSEYAN VKAQVYLAAD

251 LGAANMLKVL HPLPRIDEIA TDVDTKPHAY YFOQAGNGIF ARSALAIVVN

301 ADLAL

IIA_SEQUENCE 1.0

ID PIRB_VIBS2 STANDARD; PRT: 310 AA.

AC P96174;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUL-1998 (Rel. 38, Last annotation update)

DE ASPARTATE CARBAMOYLTRANSFERASE CATALYTIC CHAIN (EC 2.1.3.2) (ASPARTATE TRANSFERASE) (ATCACE).

GN PYR.

OS Vibrio sp. (strain 2693).

OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.

RN [1]

RP SEQUENCE FROM N.A.

RA VAN DE CASTELE M., LIANG Z., FENG Z.Y., LEGRAIN C., GLANDORFF N.; Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.

CC -1- CATALYTIC ACTIVITY: CARBAMOYL-PHOSPHATE + ASPARTATE -> ORTHOPHOSPHATE + N-CARBAMOYL-ASPARTATE.

CC -1- PATHWAY: SECOND STEP IN PYRIMIDINE BIOSYNTHESIS.

CC -1- SUBUNIT: CONTAINS SIX CATALYTIC AND SIX REGULATORY CHAINS (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE ATCASES/OTCASES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: Y09786; CAAT0923.1; -

DR HSSP: P00479; 1RAG.

DR PROSITE: PS00097; CARBAMOYLTRANSFERASE; 1.

KW Pyrimidine biosynthesis; Transferase.

SO SEQUENCE 310 AA; 34419 MW; C460D24E CRC32;

PIRB_VIBS2 Length: 310 February 14, 2000 08:02 Type: P Check: 5096 ..

1 NANPLFRKHI VSINDISRNE LELIVYIAK LKROPELHK KKVIASCFE

51 EASTRTLSFE ETAIORLGT VIGFDNASVT SLAKKGELIA DSISVSYV

101 DAFVHRHPOE GAARLASEFS NVFVINGDNG SNQPTPTIL DFIETIENG

151 CLDNLNIALV GDLKGRVTH SLQAALAKFS GCKFYFIAPD ALAMPEYICD

201 ELDEHNYSYA CYSNIEEVP EIDVLYMTRV QKRFEPETEX QHMAAGFLIS

251 ASSLKHKADN LKYLHPLPRV DEIVADVDT PYAYYQOAE NGVAREALL

301 ALVYNATIEG

IIA_SEQUENCE 1.0

ID PIRB_BACSU STANDARD; PRT: 256 AA.

AC P25983;

DT 01-MAY-1992 (Rel. 22, Created)

DT 01-MAY-1992 (Rel. 32, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE DIHYDROOrotate DEHYDROGENASE ELECTRON TRANSFER SUBUNIT.

GN PYRDI OR PYRDB.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus.

RN [1]

RP SEQUENCE FROM N.A.

RA OUTIN C.L., STEPHENSON B.T., SWITZER R.L.; "Functional organization and nucleotide sequence of the Bacillus subtilis pyrimidine biosynthetic operon.";

RL J. Biol. Chem. 266:9113-9127(1991).

RN [2]

RP FUNCTION.

RX MEDLINE: 96326349.

RA KAHNER A.E SWITZER R.L.; "Identification of a novel gene of pyrimidine nucleotide biosynthesis, pyrdii, that is required for dihydroorotate dehydrogenase activity in Bacillus subtilis.";

RL J. Bacteriol. 178:5013-5016(1996).

CC -1- FUNCTION: PROBABLE ELECTRON CARRIER PROTEIN INVOLVED IN THE TRANSFER OF REDUCING EQUIVALENTS FROM THE FLAVOPROTEIN SUBUNIT TO THE ELECTRON TRANSPORT SYSTEM IN THE CELL MEMBRANE.

CC -1- PATHWAY: FOURTH STEP IN PYRIMIDINE BIOSYNTHESIS.

CC -1- SIMILARITY: TO S.TYPHIMORIUM ASRB AND P.FURIOSUS HYDG.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: M59757; AAA21271.1; -

DR EMBL: 299112; CAB13427.1; -

DR PIR: G39845; G39845

DR SUBTILISF: HG10727; PYRDI

KW Pyrimidine biosynthesis; Electron transport.

SO SEQUENCE 256 AA; 28099 MW; 3F364D3B CRC32;

PIRB_BACSU Length: 256 February 14, 2000 08:02 Type: P Check: 2658 ..

1 MKRAYLTVCS NOQIADRVQ MWLKGELVQ ETPGQFLHL KVEAVTPLL

51 RRPISADVN FEKNEVTIY RVDEGTRLL SLKQOGEIVD VLGPNGGFP

101 VNEVOPGKTA LLVGGVGVP PLQELSKRLI EKGAVVIVL GFOSAKDVFY

151 EEECRQYGDY YVATADGSYG ETGEVVDIVK RKLEFDILL SCGTPMLKA

201 LKQEVAKREV YLSMERMG GIGACFACVC HNESETSYV KVCIDGVFK

251 AQEVAL

IIA_SEQUENCE 1.0

ID R51_DROME STANDARD; PRT: 336 AA.

AC Q27297;

DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE 15-DEC-1999 (Rel. 39, Last annotation update)
 DE DNA REPAIR PROTEIN RAD51 HOMOLOG (RECA PROTEIN HOMOLOG)
 GN RAD51 OR DMR.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 RN [1]
 RP SOURCE FROM N.A.
 RC STRAIN-CANTON-S;
 RX MEDLINE: 95161094.
 RA AKABOSHI E., INOUE Y., IYO H.;
 RT "Cloning of the cDNA and genomic DNA that correspond to the reca-like
 RT gene of Drosophila melanogaster."
 RL Jpn. J. Genet. 69:663-670(1994).
 RP [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OREGON-R;
 RX MEDLINE: 96207535.
 RA MCKEE B.D., REN X.J., HONG C.S.;
 RT "A reca-like gene in Drosophila melanogaster that is expressed at
 RT high levels in female but not male meiotic tissues."
 RL Chromosoma 104:479-488(1996).
 CC -1- FUNCTION: BINDS TO SINGLE AND DOUBLE STRANDED DNA AND EXHIBITS
 CC DNA-DEPENDENT ATPASE ACTIVITY. UNDERWINDS DUPLEX DNA (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
 CC -1- SIMILARITY: STRONG TO OTHER EUKARYOTIC RECA-LIKE PROTEIN; SOME, TO
 CC PROKARYOTIC RECA PROTEIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch)
 CC -----
 CC EMBL: D37788; BAA07039.1; -
 DR EMBL: D17725; BAA04580.1; -
 DR EMBL: L41342; AAA54873.1; -
 DR FDBASE: FB90011700; Rad51.
 KW DNA-binding; ATP-binding; Nuclear protein.
 NP-BIND 124 131
 FT SEQUENCE 336 AA: 36647 MW: 5806502 CRC32:
 SO
 RA51_DROME Length: 336 February 14, 2000 08:02 Type: P Check: 2333 ..
 1 MEKLTINVOAQ QEEEEEGL SVTKLIGSI TANDIKLQO ASLITVESVA
 51 NATKQIMAI PGIGGKVEO IITENKLVLP LGFLSARTEY OMRADEVOLS
 101 TGSKELDKLL GGGIETGSIY EIFGEFRCKG TOLCHTLAVY COLPISSKGG
 151 EGKMCYIDTE NTFPERLAA IAQRKLNES EVDNVAFTF AHSNDQOTKL
 201 IOMAGMLFE SRYALIVDS AMALYRSDYI GREGELARON HGLFLRMILQ
 251 RLADFEQAV VITNOYASL DGAPGMFAD KPIGIGHMAH SSTRLYLK
 301 GKGEIRICKI YDSPCLPESF AMFALLPGCI GDARES
 11AA_SEQUENCE 1.0
 ID RBRC_CHRV1 STANDARD; PRT; 302 AA.
 AC P25544;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 01-MAY-1992 (Rel. 22, Last annotation update)
 DE RUBISCO OPERON TRANSCRIPTIONAL REGULATOR.
 DE RBRC.
 GN Chromatium vinosum.

OC Bacteria; Proteobacteria; gamma subdivision; Chromatiaceae;
 OC Allochromatium.
 RN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE: 91317745.
 RA VIALE A.M., KOBAYASHI H., AKAZAWA T., HENIKOFF S.;
 RT "rbcr, a gene coding for a member of the lysr family of
 RT transcriptional regulators, is located upstream of the expressed set
 RT of ribulose 1,5-bisphosphate carboxylase genes in the
 RT photosynthetic bacterium Chromatium vinosum."
 RL J. Bacteriol. 173:5224-5229(1991).
 RN [2]
 RN SEQUENCE OF 1-57 FROM N.A.
 RP MEDLINE: 89213919.
 RX VIALE A.M., KOBAYASHI H., AKAZAWA T.;
 RA "Expressed genes for plant-type ribulose 1,5-bisphosphate
 RT carboxylase/oxygenase in the photosynthetic bacterium Chromatium
 RT vinosum, which possesses two complete sets of the genes."
 RL J. Bacteriol. 171:2391-2400(1989).
 CC -1- FUNCTION: TRANS-ACTING TRANSCRIPTIONAL REGULATOR OF RUBISCO
 CC -1- GENES (RECB) EXPRESSION.
 CC -1- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
 CC REGULATORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch)
 CC -----
 CC EMBL: M64032; AAA23327.1; -
 DR EMBL: M26396; -; NOT_ANNOTATED_CDS.
 DR PIR: A40369; A40369.
 DR PROSITE: PS00044; HTR_LYSR_FAMILY. 1.
 DR PIRAM: PF00126; HTR.1; 1.
 KW Transcription regulation; Activator; DNA-binding.
 FT DNA BIND 20 39
 FT H-T-H MOTIF (BY SIMILARITY).
 FT SEQUENCE 302 AA: 34092 MW: CA38CF20 CRC32:
 SO
 RBRC_CHRV1 Length: 302 February 14, 2000 08:02 Type: P Check: 4295 ..
 1 MHVSLQNAV PEAVARHSY TRAAEHLHS QPAVEMOYRQ LEDIGLSLF
 51 ERGKQVLT ENGREVPHYS RAIGQSIREM EEVLSLKGV SRGSLRIAVA
 101 STVTFAPRL NAIFQRRSG IGLRLDYTNR ESYVOMDSN SVDLYMGVP
 151 PRNVEVEADA FMDNPLVIA PPDHPLAGER AISLARLAE TFWRECGSG
 201 TROMERFES ERGQIRHGM QMTRNEAVKO AVRSGLGLSV VSLHTILEL
 251 ETRRLVTLDV EGFPDRQWY LVYRGRKRLS PAAGAFRETV LSEARMHCR
 301 LG
 11AA_SEQUENCE 1.0
 ID RBRC_SOLTU STANDARD; PRT; 181 AA.
 AC P10647;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL CHAIN C PRECURSOR
 DE (EC 4.1.1.39) (RUBISCO SMALL SUBUNIT C).
 GN RBRC-C.
 OS Solanum tuberosum (Potato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;
 OC Solanum.
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-CV. HH1201/7;
 RX MEDLINE: 88124937;
 RA WOLTER F.P., FRITZ C.C., WILMITZER L., SCHELL J., SCHREIER P.H.;
 RT "rbcs genes in Solanum tuberosum: conservation of transit peptide and
 RT exon shuffling during evolution.";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:846-850(1988).
 CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
 CC D-RIBULOSE 1,5-BISPHOSPHATE. THE PRIMARY EVENT IN PHOTOSYNTHETIC
 CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
 CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATORY PROCESS. BOTH
 CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
 CC ACTIVE SITE.
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) =
 CC 2 3-PHOSPHO-D-GLYCERATE.
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) =
 CC 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.
 CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.
 CC -1- SIMILARITY: BELONGS TO THE RUBISCO SMALL CHAIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: J03613; AAA33838.1; -
 CC PIR: A31083; RKPOSC.
 CC DR HSSP: P00866; 1RLC.
 CC DR MENDEL: 15074; SOLTU:rbcs:tm15074.
 CC PRFAM: PF00101; RUBISCO_small1.1
 CC KW Photosynthesis; Carbon dioxide fixation; Photorespiration;
 CC Lyase; Oxidoreductase; Monooxygenase; Chloroplast; Transit peptide;
 CC Multi-gene family.
 CC FT TRANSIT 58 CHLOROPLAST (BY SIMILARITY).
 CC CHAIN 181 RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL
 CC SEQUENCE 181 AA; 20368 MW; 44574D48 CRC32;
 SO RBSO_SOLTU Length: 181 February 14, 2000 08:02 Type: P Check: 4699 ..
 1 MASSIVSSAA VATRSNVAQA SWAPFTGLK SAASFVTKK NNNVDITSLA
 51 SNGGRVRCMQ VMPPIKMKY ETLSTLPDLS DEQLLEVEY LKNGWVPC
 101 EEFTEHGVY RHNSSPGY DGRYWTMVKL PMFGCTDQTO VLAEOEAKN
 151 AVPAWIRII GFDNVRQVOC ISFIAYKPEG Y
 11AA_SEQUENCE 1.0 STANDARD; PRT; 181 AA.
 ID RBS1_LYCES
 AC P08706; STANDARD; PRT; 181 AA.
 DR 01-JAN-1988 (Rel. 06, Created)
 DR 01-NOV-1988 (Rel. 16, Last sequence update)
 DR 15-JUL-1999 (Rel. 38, Last annotation update)
 DE RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL CHAIN 1 PRECURSOR
 DE (EC 4.1.1.39) (RUBISCO SMALL SUBUNIT 1) (LESS17).
 GN RBGS-1.
 OS Lycopersicon esculentum (Tomato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Euphylliphytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Asteridae; easterids I; Solanales; Solanaceae;
 OC Solanum.
 RN [1]
 RP SEQUENCE FROM N.A. (LESS 17).
 RC STRAIN-CV. VF36;
 RX MEDLINE: 87163513.
 RA MCKNIGHT T.D., ALEXANDER D.C., BABCOCK M.S., SIMPSON R.B.;
 RT "Nucleotide sequence and molecular evolution of two tomato genes
 RT encoding the small subunit of ribulose-1,5-bisphosphate
 RT carboxylase.";

RL Gene 48:23-32(1986).
 RN [2]
 RN SEQUENCE FROM N.A. (RBGS-1).
 RP MEDLINE: 86233336.
 RX PICHERSKY E., BERNATZKY R., TANKSLEY S.D., CASHMORE A.R.;
 RA "Evidence for selection as a mechanism in the concerted evolution of
 RA Lycopersicon esculentum (tomato) genes encoding the small subunit of
 RA ribulose-1,5-bisphosphate carboxylase/oxygenase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:3880-3884(1986).
 RN [3]
 RP SEQUENCE FROM N.A. (RBGS-1).
 RC STRAIN-CV. VENT CHERRY LA1221;
 RX MEDLINE: 88038372.
 RA SUGITA M., MANZARA T., PICHERSKY E., CASHMORE A., GRUSSEMAN W.;
 RT "Genomic organization, sequence analysis and expression of all five
 RT genes encoding the small subunit of ribulose-1,5-bisphosphate
 RT carboxylase/oxygenase from tomato.";
 RL Mol. Gen. Genet. 209:247-256(1987).
 RN [4]
 RP SEQUENCE OF 1-20 FROM N.A. (RBGS-1).
 RC STRAIN-CV. VENT CHERRY LA1221; TISSUE=ROOT;
 RX MEDLINE: 93144693.
 RA MANZARA T., CARRASCO P., GRUSSEMAN W.;
 RT "Developmental and organ-specific changes in DNA-protein interactions
 RT in the tomato rbcs1, rbcs2 and rbcs3a promoter regions.";
 RL Plant Mol. Biol. 21:69-88(1993).
 CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
 CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
 CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
 CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATORY PROCESS. BOTH
 CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
 CC ACTIVE SITE.
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) =
 CC 2 3-PHOSPHO-D-GLYCERATE.
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) =
 CC 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.
 CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.
 CC -1- SIMILARITY: BELONGS TO THE RUBISCO SMALL CHAIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M15235; AAA34191.1; -
 CC EMBL: M15542; AAA34188.1; -
 CC EMBL: X05982; CAA29400.1; -
 CC EMBL: X66068; CAA46868.1; -
 CC PIR: S02364; RKTOS1.
 CC DR HSSP: P00866; 1RLC.
 CC DR PRFAM: PF00101; RUBISCO_small1.
 CC KW Photosynthesis; Carbon dioxide fixation; Photorespiration;
 CC Lyase; Oxidoreductase; Monooxygenase; Chloroplast; Transit peptide;
 CC Multi-gene family.
 CC FT TRANSIT 58 CHLOROPLAST (BY SIMILARITY).
 CC CHAIN 181 RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL
 CC SEQUENCE 181 AA; 20307 MW; 981E9B5C CRC32;
 SO RBS1_LYCES Length: 181 February 14, 2000 08:02 Type: P Check: 5240 ..
 1 MASSIVSSAA AATRSNVAQA SWAPFTGLK SAASFVTKK NNNVDITSLA
 51 SNGGRVRCMQ VMPPIKMKY ETLSTLPDLS DEQLLEVEY LKNGWVPC
 101 EEFTEHGVY RHNSSPGY DGRYWTMVKL PMFGCTDQTO VLAEOEAKK
 151 AVPAWIRII GFDNVRQVOC ISFIAYKPEG F

11AA_SEQUENCE 1.0 STANDARD; PRT: 180 AA.
 ID RB51_PETSP
 AC P04714.1987 (Rel. 05, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL CHAIN SSUB8 PRECURSOR
 DE (EC 4.1.1.39) (RUBISCO SMALL SUBUNIT SSUB8).
 GN RB51.
 OS Petunia sp. (Petunia).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;
 OC Petunia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. MITCHELL;
 RX MEDLINE; 86205237.
 RA TURNER N.E., CLARK W.G., TAVOR G.J., HIRONAKA C.M., FRALEY R.T.,
 RA SHAH D.M.;
 RT "The genes encoding the small subunit of ribulose-1,5-bisphosphate
 RT carboxylase are expressed differentially in petunia leaves."
 RT Nucleic Acids Res. 14:3325-3342(1986).
 CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
 CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
 CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
 CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
 CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
 CC ACTIVE SITE.
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) =
 CC 2,3-PHOSPHO-D-GLYCERATE.
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) =
 CC 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGYCOLATE.
 CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.
 CC -1- SIMILARITY: BELONGS TO THE RUBISCO SMALL CHAIN FAMILY.
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X03820; CAA27444.1; -
 CC PIR: A24917; RKPS8.
 CC HSSP: P00866; 1RJC.
 CC MENDEL: 15078; PETSP; rbcS; tm15078.
 CC DR PFAM: PF00101; RUBISCO_small; 1.
 CC DR PHOSYNTHESIS: Carbon dioxide fixation; Photorespiration;
 CC KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast; Transit peptide;
 CC Multigene family.
 CC FT TRANSIT 1 57 CHLOROPLAST (BY SIMILARITY).
 CC FT CHAIN 58 180 RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL
 CC FT CHAIN SSUB8.
 CC SO SEQUENCE 180 AA; 20370 MW; 71E5C2F3 CRC32;
 RB51_PETSP Length: 180 February 14, 2000 08:02 Type: P Check: 2967 ..

01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL CHAIN 1 PRECURSOR
 DE (EC 4.1.1.39) (RUBISCO SMALL SUBUNIT 1).
 GN RB51.
 OS Solanum tuberosum (Potato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;
 OC Solanum.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-HH1201/7;
 RX MEDLINE; 8812437.
 RA WOLTER F.P., FRITZ C.C., WILMITZER L., SCHELL J., SCHREIER P.H.;
 RT "rbcS genes in Solanum tuberosum: conservation of transit peptide and
 RT exon shuffling during evolution."
 RT Proc. Natl. Acad. Sci. U.S.A. 85:846-850(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. AM 80.5793;
 RA FRITZ C.C., WOLTER F.P., SCHENKMEYER V., HERGET T., SCHREIER P.H.;
 RT Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
 CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
 CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
 CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
 CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
 CC ACTIVE SITE.
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) =
 CC 2,3-PHOSPHO-D-GLYCERATE
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) =
 CC 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGYCOLATE.
 CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.
 CC -1- SIMILARITY: BELONGS TO THE RUBISCO SMALL CHAIN FAMILY.
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X69759; CAA49413.1; -
 CC PIR: S31083; RKPOS1.
 CC DR PIR: S31497; S31497.
 CC DR HSSP: P00866; 1RJC.
 CC MENDEL: 490; SOUTU; rbcS; 1.
 CC DR PFAM: PF00101; RUBISCO_small; 1.
 CC DR PHOSYNTHESIS: Carbon dioxide fixation; Photorespiration;
 CC KW Lyase; Oxidoreductase; Monooxygenase; Chloroplast; Transit peptide;
 CC Multigene family.
 CC FT TRANSIT 1 58 CHLOROPLAST (BY SIMILARITY).
 CC FT CHAIN 59 181 RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL
 CC FT CHAIN 1.
 CC SO SEQUENCE 181 AA; 20556 MW; 5EA0260B CRC32;
 RB51_SOUTU Length: 181 February 14, 2000 08:02 Type: P Check: 4863 ..

11AA_SEQUENCE 1.0 STANDARD; PRT: 181 AA.
 ID RB51_SOUTU
 AC P26574.1
 DT 01-AUG-1992 (Rel. 23, Created)

11AA_SEQUENCE 1.0 STANDARD; PRT: 180 AA.
 ID RB52_LYCES
 AC P07179.1
 DT 01-APR-1988 (Rel. 07, Created)

DT 01-NOV-1990 (Rel. 16, last sequence update)
 DT 15-JUL-1999 (Rel. 38, last annotation update)
 DE RUBULOSE BIPHOSPHATE CARBOXYLASE SMALL CHAIN 2A PRECURSOR
 DE (EC 4.1.1.39) (RUBISCO SMALL SUBUNIT 2A) (LESS 5).
 GN RBGS-2A.
 OS Lycopersicon esculentum (Tomato).
 CC Eukaryota: Viridiplantae: Streptophyta: Tracheophyta;
 CC eumhyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons:
 CC core eudicots: Asteridae: euasterids I: Solanales; Solanaceae;
 CC Solanum.
 CC [1]
 RN SEQUENCE FROM N.A. (LESS 5).
 RP STRAIN-CV: YP36.
 RX MEDLINE: 87163513.
 RA MCNIGHT T.D., ALEXANDER D.C., BABCOCK M.S., SIMPSON R.B.:
 RT "Nucleotide sequence and molecular evolution of two tomato genes
 RT encoding the small subunit of ribulose-1,5-bisphosphate
 RT carboxylase.";
 RL Gene 48:23-32(1986).
 RN [2]
 RP SEQUENCE FROM N.A. (RBGS-2A).
 RX MEDLINE: 86233336.
 RA PICHERSKY E., BERNATZKY R., TANKSLEY S.D., CASHMORE A.R.:
 RT "Evidence for selection as a mechanism in the concerted evolution of
 RT Lycopersicon esculentum (tomato) genes encoding the small subunit of
 RT ribulose-1,5-bisphosphate carboxylase/oxygenase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:3880-3884(1986).
 RN [3]
 RP SEQUENCE FROM N.A. (RBGS-2).
 RC STRAIN-CV: VENT CERRY LA1221;
 RX MEDLINE: 88038372.
 RA SUGITA M., MANZARA T., PICHERSKY E., CASHMORE A., GRUSSEM W.:
 RT "Genomic organization, sequence analysis and expression of all five
 RT genes encoding the small subunit of ribulose-1,5-bisphosphate
 RT carboxylase/oxygenase from tomato.";
 RL Mol. Gen. Genet. 209:247-256(1987).
 RN [4]
 RP REVISIONS.
 RA MANZARA T.:
 RL Submitted (AUG-1989) to the EMBL/Genbank/DBJ databases.
 RN [1]
 RP SEQUENCE OF 1-9 FROM N.A. (RBGS-2).
 RC STRAIN-CV: VENT CERRY LA1221; TISSUE-ROOT;
 RX MEDLINE: 93144693.
 RA MANZARA T., CARRASCO P., GRUSSEM W.:
 RT "Developmental and organ-specific changes in DNA-protein interactions
 RT in the tomato rbc1, rbc2 and rbc3a promoter regions.";
 RL Plant Mol. Biol. 21:69-88(1993).
 CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
 CC D-RIBULOSE 1,5-BIPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
 CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
 CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
 CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
 CC ACTIVE SITE.
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BIPHOSPHATE + CO(2) =
 CC 2,3-PHOSPHO-D-GLYCERATE.
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BIPHOSPHATE + O(2) =
 CC 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGYLICOLATE.
 CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.
 CC -1- SIMILARITY: BELONGS TO THE RUBISCO SMALL CHAIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M15236; AAA34192.1; -
 CC DR EMBL: M1543; AAA34189.1; -
 CC DR EMBL: X05983; CAA29401.2; -

DR EMBL: X66069; CAA46869.1; -
 DR PIR: S02363; RKT052.
 DR HSSP: P00866; IRLC.
 DR MENDEL: 414; LYCES:rbc2.
 DR PFAM: PF00101; RUBISCO_small.1
 KM Photosynthesis: Carbon dioxide fixation; Photorepiration;
 KM Lyase, Oxidoreductase, Monooxygenase; Chloroplast; Transit peptide;
 KM Multi-gene family.
 FT TRANSIT 1 57 CHLOROPLAST (BY SIMILARITY).
 FT CHAIN 58 180 RIBULOSE BIPHOSPHATE CARBOXYLASE SMALL
 FT VARIANT 87 87 I -> V (IN REF. 1).
 SQ SEQUENCE 180 AA: 20278 MW: 687042D CRC32:
 RBS2_LYCES Length: 180 February 14, 2000 08:02 Type: P Check: 4535 ..
 1 MASSIVISSAA VATRSNTQA SWAPFTGLK SSATPYTKR ONDITSIAS
 51 NGRVSCMQV WPPIMKRYE TSLYLPDLSL EQLSEIEYL LKNGVPCLE
 101 FETEHGFYR ENKSPGYD GRWTMVKLP MFGCTDATQV LAEYCAKKA
 151 YPOAWRIIG FDNVQYQCI SFYAKPEGY
 !!NA SEQUENCE 1.0 STANDARD; PRT; 181 AA.
 ID RBS3 SOLTU
 AC P32764;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE RUBULOSE BIPHOSPHATE CARBOXYLASE SMALL CHAIN 3 PRECURSOR
 DE (EC 4.1.1.39) (RUBISCO SMALL SUBUNIT 3).
 GN RBGS-3.
 OS Solanum tuberosum (Potato).
 CC Eukaryota: Viridiplantae: Streptophyta: Tracheophyta;
 CC eumhyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons:
 CC core eudicots: Asteridae: euasterids I; Solanales; Solanaceae;
 CC Solanum.
 CC [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN-CV: AM 80.5793;
 RX MEDLINE: 94131296.
 RA FRITZ C.C., WOLTER F.P., SCHENKMEYER V., HERGET T., SCHREIER P.H.:
 RT "The gene family encoding the ribulose-(1,5)-bisphosphate
 RT carboxylase/oxygenase (Rubisco) small subunit of potato.";
 RL Gene 137:271-274(1993).
 CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
 CC D-RIBULOSE 1,5-BIPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
 CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
 CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
 CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
 CC ACTIVE SITE.
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BIPHOSPHATE + CO(2) =
 CC 2,3-PHOSPHO-D-GLYCERATE.
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BIPHOSPHATE + O(2) =
 CC 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGYLICOLATE.
 CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.
 CC -1- SIMILARITY: BELONGS TO THE RUBISCO SMALL CHAIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X69763; CAA49417.1; -
 CC PIR: S31498; S31498.
 CC DR HSSP: P00866; IRLC.
 CC DR MENDEL: 495; SOLTU:rbc3.6.
 CC PFAM: PF00101; RUBISCO_small.1.

KW Photosynthesis: Carbon dioxide fixation; Photorespiration;
 KW Lyase: Oxidoreductase; Monooxygenase; Chloroplast; Transit peptide;
 FT Multiene family. 58 CHLOROPLAST (BY SIMILARITY).
 FT TRANSIT 1 181 RIBULOSE BIPHOSPHATE CARBOXYLASE SMALL
 FT CHAIN 59 181 CHAIN 3.
 FT SEQUENCE 181 AA; 20437 MW; 268DC949 CRC32;
 RBS3_SCHPO Length: 181 February 14, 2000 08:02 Type: P Check: 4433 ..

1 MASSIVSSAA VATRNSVAQA SNAVPTGLK SNAFPTKRN NNNVDTSLA

51 SNGGVRQMQ VWPINMKKY ETLSTLPDL DEOLKEVEY LKNGWPCPL
 101 ERETEHGFVY RENHSPGY DERYWTMML PMFGCTDQ VLAEOEAKK
 151 AYPQAMIRLI GPDVROVOC ISFIAYKPEG Y

11AA_SEQUENCE 1.0 STANDARD; PRT: 318 AA.

ID RBSK_SCHPO
 AC 060116;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE PURTIVE RIBOKINASE (EC 2.7.1.15).
 GN SPC165.02C.
 OS Schizosaccharomyces pombe (Fission yeast)
 OS Eukaryota; Fungi; Ascomycota; Archaescomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 CC Schizosaccharomycetes.
 RN SEQUENCE FROM N.A.
 RC STRAIN-972;
 RA LYNE M., WOOD V., RAJANDREAM M.A., BARRELL B.G., BECK A.,
 RA REINHARDT R.;
 RL Submitted (May-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: ATP + D-RIBOSE - ADP + D-RIBOSE 5-PHOSPHATE.
 CC -1- PATHWAY: FIRST STEP IN RIBOSE METABOLISM.
 CC -1- SIMILARITY: BELONGS TO THE PERK FAMILY OF CARBOHYDRATE KINASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AL023554; CAI19022.1;
 DR PROSITE: PS00583; PERK_KINASES_1; FALSE_NEG.
 DR PROSITE: PS00584; PERK_KINASES_2; 1.
 DR PFM: PF00294; PKB; 1
 KW Hypothetical protein; Transferase; Kinase
 FT SEQUENCE 318 AA; 33250 MW; D40648BF CRC32;
 RBSK_SCHPO Length: 318 February 14, 2000 08:02 Type: P Check: 3214 ..

1 MINIVLGSMTDLVMTKTI CPBGGETTNG EPDGFSTGNG GRKANQAVAV

51 ARLSNPADTK VSMGCVGD AFGVEMLSGL KKGVDVNDV KTIENKSTGV
 101 AMIIVEETGE NRILISEGAN GNDTAFVRA MEORISTCNL LIMQLEIPLE
 151 AVEIALQIAH KHGVDLNMP APAIPLSHDM ISYCAIYVFN EHEAAILNQ
 201 ADSPATLENV DAYASKLSE GYKAVIITL GSQGYVYKSA NGSALVSAC
 251 KKAADVTAA GDTFGAFSN SIHAGPGLK SLEFAKCSA ITVQRGAAS
 301 SIPSLEVDG SFNLKNT

11AA_SEQUENCE 1.0 STANDARD; PRT: 180 AA.

ID RBS_MUSAC

AC 024045;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE RIBULOSE BIPHOSPHATE CARBOXYLASE SMALL CHAIN PRECURSOR (EC 4.1.1.39)
 DE (RUBISCO SMALL SUBUNIT).
 GN RBCS1.
 OS Musa acuminata (Banana)
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Euphylliphytes; Spermatophyta; Magnoliophyta; Liliopsida;
 CC Zingiberales; Musaceae; Musa.
 RN [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE-LEAF;
 RA REGEV I., KHAYAT E., GERSTEIN S.;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
 CC D-RIBULOSE 1,5-BIPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
 CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
 CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
 CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
 CC ACTIVE SITE.
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BIPHOSPHATE + CO(2) =
 CC 2-3-PHOSPHO-D-GLYCERATE.
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BIPHOSPHATE + O(2) =
 CC 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.
 CC -1- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.
 CC -1- SIMILARITY: BELONGS TO THE RUBISCO SMALL CHAIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF008214; BAB63287.1;
 DR PFM: PF00101; RUBISCO_small; 1.
 KW Photosynthesis; Carbon dioxide fixation; Photorespiration;
 KW Lyase: Oxidoreductase; Monooxygenase; Chloroplast; Transit peptide.
 FT TRANSIT 1 180 CHLOROPLAST (BY SIMILARITY).
 FT CHAIN 59 180 RIBULOSE BIPHOSPHATE CARBOXYLASE SMALL
 FT SEQUENCE 180 AA; 20511 MW; 3BB9BCBD CRC32;
 RBS_MUSAC Length: 180 February 14, 2000 08:02 Type: P Check: 2518 ..

1 MVSSMNVSSA ATFTASPAQ SSWAPFTGL KASAFVPR KPNADLSHP

51 SNGRVOQCMK VPIEGVKKF ETLSTLPTR DEALVKQLEY LRSKWLPCL
 101 EFCFKGVYR EHHRSBGYD GRVWTMMLP MGCCTDAVOV AKVEBECKE
 151 YPFAFIRITG FDNKROVOCI SFIAKPTGY

11AA_SEQUENCE 1.0 STANDARD; PRT: 298 AA.

ID REGS_DROME
 AC 094913;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE RHYTHMICALLY EXPRESSED GENE 5 PROTEIN (DREG-5).
 GN REG-5.
 OS Drosophila melanogaster (Fruit fly).
 OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-CANTON-S; TISSUE-HEAD;
 RX MEDLINE: 96203080.

RA VAN GELDER R.N., KRASNOV M.A.;
 RT "A novel circadianly expressed *Drosophila* melanogaster gene dependent
 RT on the period gene for its rhythmic expression.";
 RL EMBL J. 15:1625-1631(1996).
 CC -1- FUNCTION: INVOLVED IN THE GENERATION OF BIOLOGICAL RHYTHMS
 CC (POTENTIAL). IN THE HEAD, OSCILLATES IN ABUNDANCE WITH A DAILY
 CC PEAK DURING EARLY NIGHT, EVEN UNDER CONSTANT DARKNESS. OSCILLATION
 CC IS DEPENDENT ON PERIOD (PER) FUNCTION.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN HEAD, BUT NOT IN THE BODY.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN 24 HOURS EMBRYO.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; 065103; AAC47267.1; -
 CC DR EMBASE; FB90015801; Reg-5.
 CC KW Biological rhythms.
 CC FT DOMAIN 22 25 POLY-SER.
 CC SQ SEQUENCE 298 AA; 32202 MW; F901A2C6 CRC32;
 REGS_DROME Length: 298 February 14, 2000 08:02 Type: P Check: 7903 ..
 1 MTTAKVILA CCLGAFHIQ ISSSSAIPIM EFLTRNEKMS HLYSTFAQV
 51 SVHCKSTAAV GGLPVNCKH NLIGYSARKL QTLSDVQLEA LDPYORDANE
 101 LIMSIMSDH PGASLVYTR OPLQPLPTP PASSLILTR QQLPGASHA
 151 HPQSSGSAI NPFESEGEK HKYAMDMDKA YGYGQSSSE LVYAALNSE
 201 PSKRRLTGPL VIRVRPDGSP VEEDKMPLP RDEDLPLTSS WGRSQAQA
 251 PQRNRNOLK AALRLHPAE RPARPPPAE ASVPAPGVR SASEPQA
 11AA_SEQUENCE 1.0
 ID REHY_TORRU STANDARD; PRT; 218 AA.
 AC P52574;
 DT 01-OCT-1986 (Rel. 34, Created)
 DT 01-OCT-1986 (Rel. 34, Last sequence update)
 DT 13-DEC-1998 (Rel. 37, Last annotation update)
 DE REHYDRIN.
 OS Tortula ruralis (Moss).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryopsida;
 CC Bryidae; Pottiaceae; Tortula.
 CC [1]
 CC SEQUENCE FROM N.A.
 CC STRAIN-STAR MOSS; TISSUE-LEAF;
 CC OLIVER M.J., SCOTT H.B. II;
 CC "Desiccation-tolerance and gene expression: Analysis of a recovery
 CC clone, Tr288, and its implications in mRNA storage during drying.";
 CC J. Exp. Bot. 45:577-583(1994).
 CC -1- FUNCTION: ASSOCIATED WITH THE REHYDRATION EVENTS INVOLVED IN THE
 CC RECOVERY OF THE DESICCATION-TOLERANT MOSS.
 CC -1- SIMILARITY: BELONGS TO THE AHPC/TSA FAMILY. REHYDRIN SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; 040818; AAA83758.1; -
 CC DR HSSP; P30041; IPRX.
 CC DR PFAM; PF00578; AHPC-TSA; 1.
 CC KW Antioxidant.
 CC SQ SEQUENCE 218 AA; 24084 MW; ADPD8F42 CRC32;

REHY_TORRU Length: 218 February 14, 2000 08:02 Type: P Check: 4850 ..
 1 MGGWALGDL VPDQADSTM GHKVDYCK DGMITFSHP GDVPCVCTTE
 51 LGKIAVNEE PEKRVKLLG LSTIDYEDNO GMINDESYT PDAVLYKPL
 101 ADDRKRTYA LNMADPDCKD ANKPPLASA LHIISPCRL KSLILPGTT
 151 GRNFDVLRV LDSLQASKH KATPANWCK GEPVVISPV SDEKAMRFP
 201 QGMETVNLKR ALRMTFVD
 11AA_SEQUENCE 1.0
 ID REP2_YEAST STANDARD; PRT; 296 AA.
 AC P03872;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 21-NOV-1995 (Rel. 32, Last annotation update)
 DE TRANS-ACTING FACTOR C (REP2) (PROTEIN CHARLIE).
 GN REP2.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Plasmid 2-micron.
 CC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
 CC Saccharomycetaceae; Saccharomyces.
 CC [1]
 CC SEQUENCE FROM N.A.
 CC STRAIN-A364A D5;
 CC MEDLINE; 81012161.
 CC RA HARTLEY J.L., DOWNSON J.F.;
 CC "Nucleotide sequence of the yeast plasmid.";
 CC Nature 286:860-864(1980).
 CC -1- FUNCTION: PLASMID PARTITION REQUIRE REP1, REP2, AND A CIS-ACTING
 CC DNA SEQUENCE (KNOWN AS STB).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; J01347; AAB59343.1; -
 CC DR EMBL; V01323; CA24633.1; -
 CC DR PIR; A04504; PDHYC.
 CC DR SGD; L0001612; REP2.
 CC KW Plasmid; Trans-acting factor.
 CC SQ SEQUENCE 296 AA; 33196 MW; E280A192 CRC32;
 REP2_YEAST Length: 296 February 14, 2000 08:02 Type: P Check: 6596 ..
 1 MODIEFANKL TVKARTAVS MDVCLFTDM IAPVDIDIE SKRKDELLF
 51 PEYVIRPMES LITGRPGLD SAAEDSSVSS DSAEVLPLA AMWVERDS
 101 TENGWLSOE ASQAALDML QNNKLIDNRK QLYKSIATII GLPEKDKR
 151 ATEMLRKKMD CTQLLVPPAP TEEDVAKLVS VTQLLTLYV PDROAALIGD
 201 LEPIESLKI FNSFNEIAE NRLQKSEL EGTEVNHAN TNEEVSRT
 251 RSRDINARGA YKLONTTTEG PRAVPTKKR VATRVGRKS RNTSRV
 11AA_SEQUENCE 1.0
 ID REFB_NEIGO STANDARD; PRT; 346 AA.
 AC P37761;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE DTP-GLUCOSE 4,6-DEHYDRATASE (EC 4.2.1.46).
 GN REFB.
 OS Neisseria gonorrhoeae.

CC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 RN [1]
 CC SEQUENCE FROM N.A.
 RP SRRAIN-MS11.
 RX MEDLINE: 95050260.
 RA ROBERTSON B.D., FROSCH M., VAN PUTTEN J.P.M.:
 RT "The identification of cryptic rhamnose biosynthesis genes in
 CC Neisseria gonorrhoeae and their relationship to lipopolysaccharide
 CC biosynthesis." 176:6915-6920(1994).
 RL J. Bacteriol. 176:6915-6920(1994).
 CC -1- CATALYTIC ACTIVITY: DTD-GLUCOSE = DTD-4-DEHYDRO-6-DEOXY-D-
 CC GLUCOSE + H(2O).
 CC -1- COFACTOR: NAD.
 CC -1- PATHWAY: DTD-4-RHAMNOSE BIOSYNTHESIS WITHIN THE O ANTIGEN
 CC BIOSYNTHESIS PATHWAY OF LIPOPOLYSACCHARIDE BIOSYNTHESIS.
 CC -1- SIMILARITY: STRONG, TO OTHER DTD-GLUCOSE 4,6-DEHYDRATASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z33742; CAA83652.1; -
 CC EMBL: Z21508; CAA79718.1; -
 CC PIR: S47045; S47045.
 CC PRF: PFO1370; EPImerase; 1.
 CC K M Lipopolysaccharide biosynthesis; lyase; NAD.
 CC N P BIND 13 19
 CC SEQUENCE 346 AA; 38828 MW; FB9FD6A3 CRC32;
 REBB_LENGTH: 346 February 14, 2000 08:02 Type: P Check: 5086 ..
 1 MOTGKKNIL VTGAGGFIGS AVYRHIIONT RDSVYNIDKL TYAGNLESTIT
 51 DIADNPYAF EYDIDCDRAE LDRYFAOYRP DAVMHLAES HYDRAIGSAG
 101 EFRITNIVGT FDLLEAARAY WOOMPSEKRE AFREHNIHSTD EYVGLHGTD
 151 DLFETTPYA PSSPYASAKA ADHLVYRAMQ RYRLSPISV NCSNNGPQRO
 201 FPEKILPMI LNALSGKPLP VYDGAQOIRD WLEVEDHARA LYQVTEGVY
 251 GETYNIGGN EKTNLEVVKT ICALLEELAP EKPAVARE DLITVQDRP
 301 GHDAVAYDA AKTRDGLMW PLTFESGLR KTYQWYLDNK TRQNA
 11AA_SEQUENCE 1.0
 ID_RHE_MACMU STANDARD; PRT; 416 AA.
 AC Q28849;
 DI 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE RH-LIKE PROTEIN (RHESUS-LIKE PROTEIN).
 OS Macaca mulatta (Rhesus macaque).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae;
 CC Macaca.
 CC [1]
 RN SEQUENCE FROM N.A.
 RP TISSUE-BONE MARROW;
 RX MEDLINE: 94223693.
 RA MOURO I., LE VAN KIM C., CHERIF-ZAHAR B., SALVIGNOL I., BLANCHER A.,
 RA CARTON J.-P., COLIN Y.:
 RT "Molecular characterization of the Rh-like locus and gene transcripts
 CC from the rhesus monkey (Macaca mulatta)."
 RL J. Mol. Evol. 38:169-176(1994).
 CC -1- FUNCTION: MAY BE PART OF AN OLIGOMERIC COMPLEX WHICH IS LIKELY TO
 CC HAVE A TRANSPORT OR CHANNEL FUNCTION IN THE ERYTHROCYTE MEMBRANE.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE RH FAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: S70343; AAB30637.1; -
 CC PRF: PFO0909; Ammonium transp. 1.
 CC K M Erythrocyte; Transmembrane; Polymorphism.
 CC FT INITIATOR 0
 CC FT TRANSMEM 11
 CC FT TRANSMEM 43
 CC FT TRANSMEM 76
 CC FT TRANSMEM 124
 CC FT TRANSMEM 171
 CC FT TRANSMEM 202
 CC FT TRANSMEM 237
 CC FT TRANSMEM 264
 CC FT TRANSMEM 286
 CC FT TRANSMEM 330
 CC FT TRANSMEM 357
 CC FT VARIANT 1
 CC FT VARIANT 70
 CC FT VARIANT 106
 CC FT VARIANT 191
 CC FT VARIANT 261
 CC FT VARIANT 342
 CC SEQUENCE 416 AA; 45672 MW; 6AAEB87E CRC32;
 RHL_MACMU Length: 416 February 14, 2000 08:02 Type: P Check: 4371 ..
 1 SKRPSVNC CLPLMALTE ALILLFFET TYDASLSDQ KGLVASYQVC
 51 QDLTVNAVIG LGFTSNLR NSMSVAENL FLALGVOMA ILDFGLSQF
 101 SPGRVAIKLE SIKLATRSTW SMLSANAVL GAVNIYOLVY MELVELTVFG
 151 TMRIVINIF KIDYGNMMH IHVPAVFGI TVAMCLPKPL PKGTEDIQOT
 201 TSPSLFAML GTFLMFWP TENSALLNP IERKNAVST YYALAVSAVT
 251 AISVSLAPR GKKIMTYM NALAGVAL SASCHVISP WIAMVLGLVA
 301 GLISIGAKC LPVCENRYLG IHESHVHYT FGLPALLGEI TYIVMALRV
 351 VMASNNWIG QVLLSTGLS LAMAMSTISG LITGLILNK IMKGPVAKY
 401 FDDQAFWEFP HLAVGF
 11AA_SEQUENCE 1.0
 ID_RHO2_HUMAN STANDARD; PRT; 227 AA.
 AC P52158; O99535; O00680; O00734.
 DI 01-OCT-1996 (Rel. 34, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE RHO-RELATED GTP-BINDING PROTEIN RHO7.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Primates; Catarrhini; Homidae; Homo.
 CC [1]
 RN SEQUENCE FROM N.A.
 RP TISSUE-BRAIN;
 RA CHARDIN P.:
 RT Submitted (FEB-1996) to the EMBL/Genbank/DBJ databases.
 RL [2]
 RN SEQUENCE FROM N.A.
 RP MEDLINE: 97092865.
 RA SMITH T.M., LEE M.K., SZABO C.I., JEROME N., MCEUEN M., TAYLOR M.,
 RA HOOD L., KING M.-C.;

RT "Complete genomic sequence and analysis of 117 kb of human DNA
RT containing the gene BRCA1."
RL Genome Res. 6:1029-1049(1996).
CC -1- SIMILARITY: BELONGS TO THE RHO FAMILY IN THE RAS SUPERFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X85456; CAA64726.1; -
DR EMBL: L78833; AAC37595.1; -
DR HSSP: P06749; IABZ.
DR MTM: 601555; -
DR PFAM: PF00071; ras; 1.
DR GTP-binding; Prenylation; Lipoprotein.
KW NP_BIND 14 21 GTP (BY SIMILARITY).
FT NP_BIND 61 65 GTP (BY SIMILARITY).
FT NP_BIND 119 122 GTP (BY SIMILARITY).
FT DOMAIN 36 44 EFFECTOR REGION (POTENTIAL).
FT LIPID 224 GERANYL-GERANYL (BY SIMILARITY).
SQ SEQUENCE 227 AA; 25369 MW; 98964266 CRC32;
RH07_HUMAN Length: 227 February 14, 2000 08:02 Type: P Check: 2901 ..
1 MEGOSGRCKI VVGDGDEGK TALLQFARK AVPGSVPTV FENTASFEI
51 DKRRILNMW DSSSSSYDN VRPLAYRPSD AVLCFEDISR PETLDSLAK
101 WQETGDFCP NAKVYVGCCK LDMRTDLATL RELSGRRLP VTHQGVYLA
151 KQVNGSYIYE CSSRSERSY RQVPHVATVA SLGRGRQLR RTDSRRMQR
201 SAQLSGRPDR GNEGIRHNR AKSCNLM
11AA_SEQUENCE 1.0
ID RIB7_ARCFU STANDARD; PRT; 219 AA.
AC 028272;
DT 15-DEC-1999 (Rel. 39, Created)
DT 15-DEC-1999 (Rel. 39, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE PUTATIVE 5-AMINO-6-(5-PHOSPHORIBOSYLAMINO)URACIL REDUCTASE
DE (EC 1.1.1.193) (HTP REDUCTASE).
GN AF2007.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
CC [1]
RN SEQUENCE FROM N.A.
RC STRAIN-VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE: 98049343.
RA KLENN H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA KLENN K.A., DOBSON R.J., GWINN M., HICKER E.K., PETERSON J.D.,
RA RICHARDSON D., KERVINGE A.R., GRAM M.E., KIRKPIDS N.C.,
RA FLEISCHMANN R.D., QUDKEMBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA PETERSON S., REICH C.I., MCNEILL L.K., BADGER J.H., GLODEK A., ZHOU L.,
RA OVERBEER R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTERBACK T.,
RA COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,
RA SADDON P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA VASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
RA VENTER J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -1- CATALYTIC ACTIVITY: 5-AMINO-6-(5-PHOSPHORIBOSYLAMINO)URACIL +
CC NAD(+) -> 5-AMINO-6-(5-PHOSPHORIBOSYLAMINO)URACIL + NADPH.
CC -1- PATHWAY: THIRD STEP OF RIBOFLAVIN BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE HTP REDUCTASE FAMILY.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE00064; AAB89247.1; -
DR TIGR: AF2007.
KW Hypothetical protein; Riboflavin biosynthesis; Oxidoreductase; NADP.
SQ SEQUENCE 219 AA; 24578 MW; 7DABR807 CRC32;
RIB7_ARCFU Length: 219 February 14, 2000 08:02 Type: P Check: 4460 ..
1 MRPYFVNVV ASDGKISDE SRKQLRISCE EDLRIVDRR AESDAIMVGI
51 GTVLADPRL TVKSAELRER RQDKGEENP LRVVDSRCR VPLTARIIND
101 EARTLVAYS R IAPEREVREY KVAEVAVFG EERVELSALL EFLHKKGVRR
151 LMVEGGTLL SSLISONLV D EIRIYGPFI IGGRDSPTVC DQESFLKCR
201 IEKIRIGEG FAVTAFNR
11AA_SEQUENCE 1.0
ID RIPA_LUCFY STANDARD; PRT; 277 AA.
AC 000465;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUL-1992 (Rel. 38, Last annotation update)
DE RIBOSOME-INACTIVATING PROTEIN LUFFIN ALPHA FRECURSOR (RNA
DE N-GLYCOSIDASE) (EC 3.2.2.22).
OS Luffa cylindrica (Sponge loofah) (Sponge gourd).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Euphyllopytes; Spermatophyta; Magnoliophyta; Eudicotyledons;
OC Core eudicots; Rosidae; eurosids I; Cucurbitales; Cucurbitaceae;
OC Luffa.
CC [1]
RN SEQUENCE FROM N.A.
RC TISSUE-SEED;
RX MEDLINE: 92288316.
RA KATAOKA J., HABUKA N., MIYANO M., MASUTA C., KOIMAI A.;
RT "Nucleotide sequence of cDNA encoding alpha-luffin, a ribosome-
RT inactivating protein from Luffa cylindrica."
RL Plant Mol. Biol. 18:1199-1202(1992).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF THE N-GLYCOSIDIC BOND AT ONE
CC SPECIFIC ADENOSINE ON THE 28S RNA.
CC -1- SIMILARITY: TO OTHER BACTERIAL AND PLANTS RIBOSOME-INACTIVATING
CC PROTEINS. BELONGS TO TYPE 1 RIP.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X63371; CAA44229.1; -
DR PIR: S22494; S22494.
DR HSSP: P16094; IAHB.
DR PROSITE: PS00275; SHIGA_RICIN; 1.
DR PFAM: PF00161; RIP; 1.
KW Antiviral; Protein synthesis inhibitor; Hydrolase; Toxin; Signal.
FT SIGNAL 1 19
FT CHAIN 20 277
FT ACT SITE 179 179 RIBOSOME-INACTIVATING PROTEIN LUFFIN-
FT SEQUENCE 277 AA; 30212 MW; E0233100 CRC32;
RIPA_LUCFY Length: 277 February 14, 2000 08:02 Type: P Check: 429 ..

1 MKRFTVLILA IFVAATVEA DVRESLSGSS STYSKFTIGD LRKALPENGST
 51 VYNITLILSS ASGASRYTLM TLSNYDGKAI TAAVDVTNY IMGYLVNSTS
 101 YFENESDARK ASQYFKGST IVTLPSYGN EKLQTAAGRI REKIPLGEP
 151 LDSAITTLRH YDSTAANAAL LVITQTAEA SRFKIEGDI IERISKNOVP
 201 SLATILENE WSALSKOJOL AQINNGTFRK PYVIIDDKQ RVELTNTSK
 251 VTKNTQOLL NKONVAAP EDVSAK

11AA_SEQUENCE 1.0 STANDARD: PRT: 250 AA.
 ID RPB_LUFCY
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE RIBOSOME-INACTIVATING PROTEIN LUFCIN-B (RRNA N-GLYCOSIDASE)
 DE (BC 3.2.2.22).
 OS Lufta cylindrica (Smooth loofah) (Sponge gourd).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids I; Cucurbitales; Cucurbitaceae;
 OC Lufta.
 RN [1]
 RP SEQUENCE.
 RC TISSUE-SEED;
 RX MEDLINE: 9124848.
 RA ISLAM M.R., HIRAYAMA H., FUNATSU G.;
 RT "Complete amino acid sequence of luftin-b, a ribosome-inactivating
 RT protein from sponge gourd (Lufta cylindrica) seeds.";
 RL Agric. Biol. Chem. 55:229-238(1991).
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF THE N-GLYCOSIDIC BOND AT ONE
 CC SPECIFIC AMINOACID ON THE 28S RRNA.
 CC -1- SIMILARITY: NO OTHER BACTERIAL AND PLANTS RIBOSOME-INACTIVATING
 CC PROTEINS BELONGS TO TYPE 1 RIP.
 DR PIR: J0108; J0108.
 DR HSSP: P16094; IMOM.
 DR PROSITE: PS00275; SHIGA_RICIN: 1.
 DR PFAM: PF00161; RIP: 1.
 KW Antiviral; Protein synthesis inhibitor; Hydrolase; Toxin.
 FT ACT_SITE 160 160 BY SIMILARITY.
 FT SEQUENCE 250 AA; 27293 MW; 64A72993 CRC32;
 SQ
 RPB_LUFCY Length: 250 February 14, 2000 08:02 Type: P Check: 7681 ..

1 ANVSFSLSGA DSKYSKFT ALRKALPSKE KVSNIPLILP SASGASRYIL
 51 MOLSNDYDARA ITMAIDVTNY YIMGYLVNST SYFANESDAR LASQYFKGS
 101 TLVTIPYSGN YERLONAAKG IREKIPLGFR ALDSALTSIF HYDSTAANA
 151 FLVILQTTAE ASRFKIEGDI IERIPKNEV PSPALSLN EAMSLSKOI
 201 QLAOTNGAF RTPVIIDNK GORVEITNLA SKVQIDVNS KLLNKONIA

11AA_SEQUENCE 1.0 STANDARD: PRT: 256 AA.
 ID R21_SPIOI
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE 30S RIBOSOMAL PROTEIN L21, CHLOROPLAST PRECURSOR (CL21) (CS-L7).
 GN RPL21.
 OS Spinacia oleracea (Spinach).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Caryophyllidae; Caryophyllales; Chenopodiaceae;
 OC Spinacia.
 RN [1]
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 56-67.
 RX MEDLINE: 91168304.

RA MARTIN W., LAGRANGE T., LI Y.F., BISANT-SEYER C., MACHE R.;
 RT "Hypothesis for the evolutionary origin of the chloroplast ribosomal
 RT protein L21 of spinach.";
 RL Curr. Genet. 18:553-556(1990).
 RN [2]
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 56-75.
 RC STRAIN-CV. ALVARO;
 RX MEDLINE: 90375547.
 RA SMOOKER P.M., KRUEF V., SUBRAMANIAN A.R.;
 RT "A ribosomal protein is encoded in the chloroplast DNA in a lower
 RT plant but in the nucleus in angiosperms. Isolation of the spinach L21
 RT protein and cDNA clone with transit and an unusual repeat sequence.";
 RL Biol. Chem. 265:16699-16703(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. GEANT D'HIVER;
 RX MEDLINE: 93205007.
 RA LAGRANGE T., FRANZETTI B., AXELOS M., MACHE R., LERBS-MACHE S.;
 RT "Structure and expression of the nuclear gene coding for the
 RT chloroplast ribosomal protein L21: developmental regulation of a
 RT housekeeping gene by alternative promoters.";
 RL Mol. Cell. Biol. 13:2614-2622(1993).
 CC -1- FUNCTION: THIS PROTEIN BINDS TO 23S RIBOSOMAL RNA IN THE PRESENCE
 CC OF PROTEIN L20 (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE L21P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X56691; CAA40019.1;
 DR EMBL: M57413; AAA34041.1;
 DR EMBL: M64682; AAA74715.1;
 DR PIR: S13527; S13527.
 DR PIR: A48103; A48103.
 DR MENDEL: 10833; SPIOI; rpl21.1.
 DR PROSITE: PS01169; RIBOSOMAL_L21: 1.
 DR PFAM: PF00829; RIBOSOMAL_L21P: 1.
 KW Ribosomal protein; rRNA-binding; Chloroplast; Transit peptide.
 FT TRANSIT 1 55 CHLOROPLAST.
 FT CHAIN 56 256 50S RIBOSOMAL PROTEIN L21.
 FT DOMAIN 236 253 ALA/GLU-RICH.
 FT SEQUENCE 256 AA; 28408 MW; D1FB7F00 CRC32;
 SQ
 R21_SPIOI Length: 256 February 14, 2000 08:02 Type: P Check: 718 ..

1 MASATLAFSC SSLCATLKLIP QNLNPLLVN PLSKPFSGV VSPPSLSRLS
 51 LIPVAKRRR FQEIPEELKA EFEFFORPPN QKPQLSDVLP DDFQAREPGT
 101 PEYNDINGOF LPKKPPPPR EEIFAVVIG SRQYVIGR WIYQRLKGA
 151 TVNDKIVLNR VLVGTAKST YIGTPIVTA AVHAVNEQL LDDKIVFKY
 201 KKKKNRRNI GHROPITRIK ITGITGYEDY PASTLEAVE AKBEAAEAE
 251 AEAVPY

11AA_SEQUENCE 1.0 STANDARD: PRT: 132 AA.
 ID RRB_MARPO
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE CHLOROPLAST 30S RIBOSOMAL PROTEIN S8.
 GN RPS8.
 OS Marchantia polymorpha (Liverwort).
 OC Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiopsida;

OC Marchantiales; Marchantiaceae; Marchantia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 8906687.
 RA FUKUZAWA H., KOHCHI T., SANO T., SHIRAI H., UMESONO K., INOKUCHI H.,
 RA OZERI H., OHYAMA K.,
 RT Structure and organization of Marchantia polymorpha chloroplast
 genome. III. Gene organization of the large single copy region from
 RT tbc1 to trnI(CAU).
 RT J. Mol. Biol. 203:333-351(1988).
 RN [2]
 RP COMPLETE GENOME.
 RA OHYAMA K., FUKUZAWA H., KOHCHI T., SHIRAI H., SANO T., SANO S.,
 RA UMESONO K., SHIKI Y., TAKEUCHI M., CHANG Z., AOTA S., INOKUCHI H.,
 RT OZERI H.,
 RT Chloroplast gene organization deduced from complete sequence of
 RT liverwort Marchantia polymorpha chloroplast DNA.
 RT Nature 322:572-574(1986).
 CC -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 16S RIBOSOMAL RNA.
 CC -1- SIMILARITY: BELONGS TO THE S8P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X04465; CAA28121.1; .
 DR PIR: A02716; R3LV8.
 DR HSSP: P56209; 1SET.
 DR MENDEL: 13155; MARPO:ps8.1.
 DR PROSITE: PS00053; RIBOSOMAL_S8: 1.
 DR PFAM: PF00410; Ribosomal_S8: 1.
 KW Ribosomal protein; rRNA-binding; Chloroplast.
 SO SEQUENCE 132 AA; 14921 MW; 20331A62 CRC32;
 RRE_MARPO Length: 132 February 14, 2000 08:02 Type: P Check: 6002 ..
 1 MGNDTIANMI TSIRNANIGK IKTVQVPAIN ITRNTAKLF QEGFIDFID
 51 NKONTKDILI LNKYQGRK KSYITTLURI SKPGLIYSN HKELPKVLGG
 101 MGIVILSTSR GIMTDREARQ KRIGELLYC VW
 11AA_SEQUENCE 1.0
 ID RS13_SCHPO STANDARD; PRT; 150 AA.
 AC P28189;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE 40S RIBOSOMAL PROTEIN S13.
 GN RPS13 OR SPAC6F6.07C.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 CC Schizosaccharomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 92375702.
 RA MARKS J., SIMANIS V.,
 RT Cloning of the gene for ribosomal protein S13 from the fission yeast
 RT Schizosaccharomyces pombe.
 RT Nucleic Acids Res. 20:4094-4094(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972;
 RA GENTILES S., CHURCHER C.M., BARRELL B.G., RAJANDREAM N.A., WOOD V.,
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE S13P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X67030; CAA47424.1; .
 DR EMBL: Z69881; CAB1741.1; .
 DR PIR: S26296; S26296.
 DR PROSITE: PS00362; RIBOSOMAL_S15: 1.
 DR PFAM: PF00312; Ribosomal_S15: 1.
 KW Ribosomal protein.
 ET INIT.MET
 SO SEQUENCE 150 AA; 16822 MW; C7588E0E CRC32;
 RS13_SCHPO Length: 150 February 14, 2000 08:02 Type: P Check: 3858 ..
 1 GRMSKRGKI ASSALPYRAS PPAACKADAD SYVEQILKFS KKGSPSOIG
 51 VTLDSHGIP QVETGQKI MRILKANGLA PELPEDLYNL IKKAVSVKHK
 101 LERNRKDRDS KFFLLIESR IHLARYRK VGALPPTWKY ESATASALVA
 11AA_SEQUENCE 1.0
 ID RS3_ACHAX STANDARD; PRT; 257 AA.
 AC P41117;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)
 DE 30S RIBOSOMAL PROTEIN S3.
 GN RPS3 OR RPS3.
 OS Acholeplasma axanthum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 CC Acholeplasmataceae; Acholeplasma.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-5743;
 RX MEDLINE: 94169035.
 RA TOTH K.I., HARRISON N., SEARS B.B.,
 RT Phylogenetic relationships among members of the class Mollicutes
 RT deduced from rps3 gene sequences.
 RT Int. J. Syst. Bacteriol. 44:119-124(1994).
 CC -1- FUNCTION: THIS PROTEIN IS INVOLVED IN THE BINDING OF INITIATOR
 CC MET-tRNA (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE S3P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L22465; AAA21916.1; .
 DR PROSITE: PS00548; RIBOSOMAL_S3: 1.
 DR PFAM: PF00189; Ribosomal_S3_C: 1.
 DR PFAM: PF00417; Ribosomal_S3_N: 1.
 KW Ribosomal protein; rRNA-binding.
 SO SEQUENCE 257 AA; 28711 MW; 8035520F CRC32;
 RS3_ACHAX Length: 257 February 14, 2000 08:02 Type: P Check: 7310 ..
 1 MGRVNPDIGF RVGVIRDMDS KWDKRIKVP ALVKEDAVIR FELNKKYNNNA
 51 AVSHEIERL KELKVKRRYK ILLHSGKPGV VIGREAVATK ETIASLEKLT
 101 KKEIVFWVE VKRPEVATL VAQSMADOLE NRASFESKTK NCYAKSIKVR
 151 AKGIKTSOR LGREKARTE GYSEGQVPLH TLRADEVAT AEAQTYGIL
 201 GIKWITHEG ILPGQSHETL RKEROSSASS NNGGGRRRS RKGRRSGED

251 ATEBGN

11AA_SEQUENCE 1.0 STANDARD: PRT: 133 AA.

RS8_SYNY3

AC 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE 30S RIBOSOMAL PROTEIN S8.

GN RPSH OR RPS8 OR SL1809.

OS *Synechocystis* sp. (strain PCC 6803).

OC Bacteria; Cyanobacteria; Chroococcales; *Synechocystis*.

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE: 97061201.

RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y., MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T., HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARDO K., OKUMURA S., SHIMO S., TAKEUCHI C., MADA T., WATANABE A., YAMADA M., YASUDA M., TABATA S.;

RA "Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.";

RT DNA Res. 9:109-136(1996)

RL -1- FUNCTION: BINDS DIRECTLY TO THE CENTRAL DOMAIN OF 16S RIBOSOMAL RNA (BY SIMILARITY). BELONGS TO THE S8P FAMILY OF RIBOSOMAL PROTEINS.

CC -1- SIMILARITY: BELONGS TO THE S8P FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: D90905; BAA17336.1; -

CC DR HSSP: P56209; 1SEI.

DR PROSITE: PS00053; RIBOSOMAL_S8; 1.

DR PFM: PF00410; Ribosomal_s8; 1.

KW Ribosomal protein: rRNA-binding.

SO SEQUENCE 133 AA; 14666 MW; 3596E999 CRC32;

RS8_SYNY3 Length: 133 February 14, 2000 08:02 Type: P Check: 8926

1 MASTDISISM LTRIRNACAV RHSTQVPT KMTLSIAVL KSGFIEDS

51 ETGEGINKML VTLTKYKGT RQPLINTLQ VSKRLRYVS PSKRIRVLG

101 GIGIAIVSTS HGIMTDREAR RGIGIGELIC YIV

11AA_SEQUENCE 1.0 STANDARD: PRT: 324 AA.

RUVB_TRETH

AC 056214;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE HOLLIDAY JUNCTION DNA HELICASE RUVB.

GN RUVB.

OS *Thermus aquaticus* (subsp. *therophilus*).

OC Bacteria; *Thermus*/Deinococcus group; *Thermus* group; *Thermus*.

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE: 96196179.

RA TONG J., WETIMUR J.G.;

RA "Cloning, sequencing, and expression of *ruvB* and characterization of *ruvB* proteins from two distantly related thermophilic eubacteria.";

RT J. Bacteriol. 178:2695-2700(1996)

RL -1- FUNCTION: THE RUVB-RUVB COMPLEX IN THE PRESENCE OF ATP REMOVES CRUCIFORM STRUCTURE IN SUPERCOILED DNA WITH PALINDROMIC SEQUENCE. INDICATING THAT IT MAY PROMOTE STRAND EXCHANGE REACTIONS IN HOMOLOGOUS RECOMBINATION. RUVB IS AN HELICASE THAT MEDIATES THE

HOLLIDAY JUNCTION MIGRATION BY LOCALIZED DENATURATION AND REANNEALING (BY SIMILARITY).

CC -1- SUBUNIT: FORMS A COMPLEX WITH RUVB.

CC -1- SIMILARITY: TO OTHER ATP-BINDING DNA REPAIR PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: U22817; AAB03726.1; -

DR PFM: PF00004; AAA; 1.

DR DNA repair; SOS response; ATP-binding; DNA recombination; Helicase.

FT NP_BIND 45 52 ATP (POTENTIAL).

SO SEQUENCE 324 AA; 36015 MW; D9638BD3 CRC32;

RUVB_TRETH Length: 324 February 14, 2000 08:02 Type: P Check: 3575

1 MEDIALPRT LDEYIGOERL KOKLRVYLEA AKARKEPLEH LLFGPGGLG

51 KTLIAHYIAH ELGVNLRYVS GAIEKPGDL AAILANSLEE GDILFIDEIH

101 RLSRQAEHL YPAMEDFVMD IVIGGPAAR TIRELPERRA LIGATRPGL

151 ITATLSRFG IVEHLEYTP EELAQVMRD ARLGVIRE EALEIGRS

201 RGTNRVAKRL FRVRDFAY EGEVITRR ALEALALDL DELLEKDR

251 EILEVLLRF GAGPGVGLATL ATALSEDPGT LEEVHEPVI RQGLKRTPR

301 GRVATELAIR HGYPPVGP LLEP

11AA_SEQUENCE 1.0 STANDARD: PRT: 217 AA.

AC 527A_ECOLI

DT 01-NOV-1992 (Rel. 23, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE SIGMA CROSS-REACTING PROTEIN 27A (SCRP-27A).

GN YHBL.

OS *Escherichia coli*.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; *Escherichia*.

RN [1]

RP SEQUENCE FROM N.A.

RA STRAIN-K12 / W3110.

RA SMITLIE D.A., FUJITA N., TOWNSLEY F.M., ISHIIHAMA A., HAYWARD R.S.;

RL Submitted (xxx-1993) to the EMBL/Genbank/DBJ databases.

RC [2]

RP SEQUENCE FROM N.A.

RA STRAIN-K12 / MG1655;

RA MEDLINE: 97428617.

RA BLATTNER F.R., PLOUNKET G., III, BLOCH C.A., PERNA N.T., BURLAND V., RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F., GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J., MAU B., SHAO Y.;

RA "The complete genome sequence of *Escherichia coli* K-12.";

RT Science 277:1453-1474(1997).

RL [3]

RP SEQUENCE OF 1-24.

RA MEDLINE: 92246944.

RA UESHIMA R., FUJITA N., ISHIIHAMA A.;

RA "Identification of *Escherichia coli* proteins cross-reacting with antibodies against region 2.2 peptide of RNA polymerase sigma subunit.";

RT Biochem. Biophys. Res. Commun. 184:634-639(1992).

CC -1- FUNCTION: NOT KNOWN. CROSS-REACTS WITH ANTIBODIES AGAINST REGION 2.2 OF RPOD AND RPOH.

CC -1- SIMILARITY: BELONGS TO THE ES1 FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: D13188; BAA02487.1; -
DR EMBL: U18997; AAA58011.1; ALT-INT.
DR EMBL: AE000400; AAC76241.1; ALT-INT.
DR PIR: J0287; J0287.
DR ECGENE: EG11383; YHBL.
FT CONFLICT 111 111 C -> S (IN REF. 1).
FT CONFLICT 195 217 AONTAASGIDKLVSRVLAIE -> RRTLOKRAALISM
SQ SEQUENCE 217 AA; 22981 MW; 501BEDA CRC32;

S37A_ECOLI Length: 217 February 14, 2000 08:02 Type: P Check: 54

1 MKRIGVILSG CGVYDSEIH EAVLTLLAIS RSGAAYCFA PDKQYDVIN
51 HLTGEAMTET RNVLEIARIT TGEIRPLAQ ADAALDALI VPGEGFAAN
101 LSNFASLGE CTVDRELKAL AQAMHAGKP LGFACIAPAM LPRIEDELAR
151 LTIGTDIDTA EYLEMGAEH VPCPDIVY DEDKIVTTP AYVLAQNIAE
201 AASGIDKLVS RVYLAIE

11AA_SEQUENCE 1.0
ID SECT_RICPR STANDARD; PRT; 433 AA.

AC 094GCS5;
DT 15-DEC-1999 (Rel. 39, Created)
DT 15-DEC-1999 (Rel. 39, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE PREPROMETIN TRANSLOCASE SECT SUBUNIT.

OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.

RP SEQUENCE FROM N.A.

RC STRAIN-MADRID E;

RA MEDLINE; 99039499.

RA ANDERSSON S.G.E.; ZOMORODIPOUR A.; ANDERSSON J.O.;

RA SIEHERITZ-PONTEN T.; ALSMARK U.C.M.; PODOMSKI R.M.; NAEISLUND A.K.;

RA ERIKSSON A.-S.; WINKLER H.H.; KURLAND C.G.;

RT "The genome sequence of Rickettsia prowazekii and the origin of
mitochondria."

RL Nature 396:133-140(1998).

CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT. INTERACTS WITH SECA AND SECE
TO ALLOW THE TRANSLOCATION OF PROTEINS ACROSS THE PLASMA MEMBRANE,
BY FORMING PART OF A CHANNEL. (BY SIMILARITY).

CC -1- SUBUNIT: ONE OF SEVEN SECRETORY PROTEINS (SECA-F & SECY) THAT
COMPRISE THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS (BY
SIMILARITY).

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

CC -1- SIMILARITY: BELONGS TO THE SECT/SECT1-ALPHA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL: AJ235272; CA115079.1; -

DR PROSITE; PS00755; SECT_1; FALSE_NEG.

DR PROSITE; PS00756; SECT_2; 1.

KW Protein transport; Transmembrane; Translocation.
FT TRANSMEM 17 37 POTENTIAL.

FT TRANSMEM 71 91 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 149 169 POTENTIAL.
FT TRANSMEM 184 204 POTENTIAL.
FT TRANSMEM 212 232 POTENTIAL.
FT TRANSMEM 268 288 POTENTIAL.
FT TRANSMEM 310 330 POTENTIAL.
FT TRANSMEM 368 388 POTENTIAL.
FT TRANSMEM 388 408 POTENTIAL.
SQ SEQUENCE 433 AA; 47696 MW; 2A8BEC17 CRC32;

SECT_RICPR Length: 433 February 14, 2000 08:02 Type: P Check: 3620

1 MGONFERRKS NDVLRRIIFT LFMILICRNG SEPIPIGIDS IALNSVAERN
51 QFGILMFNM LSGSLGRMS IFALAIPIYI TASIIQILMS VAYKPLENIK
101 KEGETGRKRI NOLSRVITVL LASFOAYGVA LSLESMVTNT GPVILLAGEF
151 ERYTVITLV VGTILMLWG EQITORGIGN GTSILIFIGI ISGVPAIIS
201 MFLSRKAL SPLAIIVCI GAVLLAIITII FFEKQKRL VQPKRQVGN
251 KIYGEATHM PLKNTSGVI PPIFASILL FPTLASFSN SNSDTSMILT
301 YLGHGKPYV ILKLVILMF FSFFIYAIIV NSEETANLR KYGAYIPGR
351 PKNTSDYFD YILRLVIG GILSVICVI PELLMKXVI SLSLGSITSL
401 IVNVVLDTM TQIOTYLFSS KYEGIMKKIK LKN

11AA_SEQUENCE 1.0
ID SECT_RICPR STANDARD; PRT; 161 AA.

AC 176302; P77784;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE PHOSPHOHISTIDINE PHOSPHATASE SIXA (EC 3.1.3.3) (RX6).

GN SIXA.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

RP SEQUENCE FROM N.A.

RC STRAIN-K12 / MC4100;

RA MEDLINE; 98149313.

RA OGINO H.; MATSUBARA M.; KATO N.; NAKAMURA Y.; MIZUNO T.;

RA "An Escherichia coli protein that exhibits phosphohistidine
phosphatase activity towards the Hpt domain of the ArcB sensor
involved in the multistep His-Asp phosphorelay."

RL Mol. Microbiol. 27:573-585(1998).

CC -1- FUNCTION: EXHIBITS PHOSPHOHISTIDINE PHOSPHATASE ACTIVITY TOWARDS
THE HPT DOMAIN OF THE ARCB SENSOR INVOLVED IN THE MULTISTEP HIS-

CC -1- SUBUNIT: ONE OF SEVEN SECRETORY PROTEINS (SECA-F & SECY) THAT
COMPRISE THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS (BY
SIMILARITY).

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

CC -1- SIMILARITY: BELONGS TO THE SECT/SECT1-ALPHA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL: AJ235272; CA115079.1; -

DR PROSITE; PS00755; SECT_1; FALSE_NEG.

DR PROSITE; PS00756; SECT_2; 1.

KW Protein transport; Transmembrane; Translocation.
FT TRANSMEM 17 37 POTENTIAL.

```

CC      ASP PHOSPHORELAY.
CC      -1- SIMILARITY: BELONGS TO THE SIXA FAMILY OF PHOSPHATASES.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: D86398; BAA24878.1; -
DR      EMBL: AB000322; AAC15400.1; -
DR      EMBL: D90864; CAB22123.1; -
DR      EMBL: D90865; CAB22129.1; -
DR      EMBL: EGI4126; SIXA.
DR      ECODENSE: EGI4126; SIXA.
KW      Hydrolyase.
FT      CONFLICT 80 80 T -> K (IN REF. 3).
SQ      SEQUENCE 161 AA; 17208 MW; 396B6F03 CRC32;

SIXA_ECODI Length: 161 February 14, 2000 08:02 Type: P Check: 9360 ...

1  MGVFIMRHGD AALDAASDV RPLTNGCDE SRLMANMLKG QKVEIERVLY
51  SFLRAEOTL EEVGDCLNP SSAEVLPELT PCGDVGLVSA YLQALNTEGY
101 ASVLISHLP LVGYVALIC PGETPMFTT SALASVTIDE SGNGTNMQM
151 SPCNKKMAKA I

11AA_SEQUENCE 1.0
ID      SIXA_HAEIN STANDARD; PRT; 164 AA.
AC      P44164.
DT      01-NOV-1995 (Rel. 32, Created)
DT      01-NOV-1995 (Rel. 32, Last sequence update)
DT      15-JUL-1998 (Rel. 36, Last annotation update)
DE      PHOSPHOTRISIDINE PHOSPHATASE SIXA HOMOLOG (EC 3.1.3.-).
GN      SIXA OR H1138.
OS      Haemophilus influenzae.
OC      Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC      Haemophilus.
OC      [1]
RN      SEQUENCE FROM N.A.
RC      STRAIN-RD / KW20;
RX      MEDLINE: 95350630.
RA      FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
RA      KERAVAGE A.R., BULF C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
RA      MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
RA      SCOTT J.D., SHIPLEY R., LIU L.-I., GLODER A., KELLEY J.M.,
RA      WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,
RA      UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDER D.M., BRANDON R.C.,
RA      FINE L.D., FRITZMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,
RA      GRENY C.L., McDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
RA      VENTER J.C.,
RT      "Whole-genome random sequencing and assembly of Haemophilus
RT      influenzae Rd."
RL      Science 269:496-512(1995).
CC      -1- SIMILARITY: BELONGS TO THE SIXA FAMILY OF PHOSPHATASES.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: U32813; AAC22985.1; -
DR      TIGR: H1138; -
KW      Hydrolyase.
SQ      SEQUENCE 164 AA; 18700 MW; 3F6E5D03 CRC32;

SIXA_HAEIN Length: 164 February 14, 2000 08:02 Type: P Check: 2974 ...

```

```

1  MNIFMERGE AEYMANSKRA RHLIVYGSQ AFLQOGWLKO HSLTVINSL
51  DRIVSPYVR ADETFHOVNO AFDELEENKF EIWEGITPYG HAHSYDYLE
101 VLKDEGVKSV LIVSHPLVG EIVAEVYGRK NPISFYPAFI AQLLMQNKKS
151 EILMQASPV IYDK

11AA_SEQUENCE 1.0
ID      SPAL_SALTY STANDARD; PRT; 432 AA.
AC      P39444.
DT      01-FEB-1995 (Rel. 31, Created)
DT      01-FEB-1995 (Rel. 31, Last sequence update)
DT      01-FEB-1996 (Rel. 33, Last annotation update)
DE      PROBABLE ATP SYNTHASE SPAL/INVC (EC 3.6.1.34).
GN      SPAL OR INVC.
OS      Salmonella typhimurium.
OC      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC      Salmonella.
OC      [1]
RN      SEQUENCE FROM N.A.
RC      STRAIN-SRLT / SL1344;
RX      MEDLINE: 94321319.
RA      EICHELBERRG K., GINOCCHIO C.C., GALAN J.E.;
RT      "Molecular and functional characterization of the Salmonella
RT      typhimurium invasion genes invB and invC: homology of invC to the
RT      FOF1 ATPase family of proteins."
RT      J. Bacteriol. 176:4501-4510(1994).
[2]
RN      SEQUENCE OF 97-432 FROM N.A.
RP      MEDLINE: 94008985.
RA      GROISMAN E.A., OCHMAN H.;
RT      "Cognate gene clusters govern invasion of host epithelial cells by
RT      Salmonella typhimurium and Shigella flexneri."
RT      EMBO J. 12:3779-3787(1993).
[3]
RN      SEQUENCE OF 403-432 FROM N.A.
RP      STRAIN-SRL1 / SL1344;
RX      MEDLINE: 95272391.
RA      COLLAZO C., ZIERHNER M.K., GALAN J.E.;
RT      "Functional analysis of the Salmonella typhimurium invasion genes
RT      invI and invJ and identification of a target of the protein secretion
RT      apparatus encoded in the inv locus."
RT      Mol. Microbiol. 15:25-38(1995).
CC      -1- FUNCTION: NECESSARY FOR EFFICIENT ENTRY OF S.TYPHIMURIUM INTO
CC      CULTURED EPITHELIAL CELLS. PROBABLE CATALYTIC SUBUNIT OF A PROTEIN
CC      TRANSLOCASE. MAY ENERGIZE THE PROTEIN EXPORT APPARATUS ENCODED IN
CC      THE INV LOCUS WHICH IS REQUIRED FOR THE SURFACE PRESENTATION OF
CC      DETERMINANTS NEEDED FOR THE ENTRY OF SALMONELLA SPECIES INTO
CC      MAMMALIAN CELLS.
CC      -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC      -1- SIMILARITY: TO OTHER TRANSPORT APPARATUS ATPASES AND TO ATPASE
CC      BETA SUBUNIT.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: U08279; AAA74038.1; -
DR      EMBL: X73525; CAAS1921.1; -
DR      EMBL: U10872; AAA83429.1; -
DR      PIR: S37304; S37304.
DR      STYGENE; SG10464; SPAL.
DR      PROSITE; PS00152; ATPASE_ALPHA_BETA_1.
DR      PFAM; PF00006; ATP-synt_ad; 1.
KW      Hydrolyase; Hydrogen ion transport; ATP synthesis; ATP-binding;
KW      virulence.
SQ      NP_BIND 159 166 ATP (POTENTIAL).
FT

```

FT MUTAGEN 165 165 K->E: LOSS OF FUNCTION.
 FT CONFLICT 291 291 G->A (IN REF. 2).
 FT CONFLICT 388 401 CAKISITIGRCRG -> GENIDNRAMQMR (IN
 FT CONFLICT 407 407 R->W (IN REF. 2).
 FT CONFLICT 432 AA; 47253 MM; DC79488F CRC32:
 SPAL_SALTY Length: 432 February 14, 2000 08:02 Type: P Check: 7477

1 MKTRLLQYL AYPQKITGPI IEAELRDYAI GELECEITPWL AKRGCCCTCA
 51 GGMVLATGTHR ADAYRNCQGL SBDVLYPTG PALSAMWYGS VLGAVLDPTG
 101 KIVERFTEPV APISERVID VAPPSYASRV GVREPLITGV RAIDGLITCG
 151 VGOBNGIFAS AGCGKTMLMH MLEQTEADV FVIGLIGERG REVTEFVDM
 201 RASHKRECV LVEFISDFPS VDRCAAOIA TTVAEYFRQ GRYVLFIDS
 251 MTRARALRD VALASGERPA RRGYPASVFD NLPRLERPG GTEGSSITAF
 301 YTVLESEEE ADPMADEIRS IIDGHLYSR KLAGQGHYPA IDVLSVSRY
 351 FGQVTFTHA EQASAVRKLH TRLELOLFI DLGEYRCAK ISITIGRCNC
 401 GDSUKARLCQ PVAQYSSFD TLSCNNAFAD QN

11AA_SEQUENCE 1.0 STANDARD; PRT; 241 AA.

ID SPIR_SPIME
 AC P21625: (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 01-MAY-1991 (Rel. 18, Last annotation update)
 DE SPIRALIN.
 OS Spiroplasma melliferum.
 CC Bacteria; Firmicutes; Bacillus/Clostridium group; Molluscites;
 CC Spiroplasmataceae; Spiroplasma.
 RN [1]

RC SEQUENCE FROM N.A.
 RA STRAIN-ATCC 33219 / BC3;
 RA MEDLINE: 91008990.
 RA CHEVALIER C., SAILLARD C., BOVE J.M.;
 RT "Spiralins of Spiroplasma citri and Spiroplasma melliferum: amino
 acid sequences and putative organization in the cell membrane.";
 RL J. Bacteriol. 172:6090-6097(1990).
 CC -1- SUBUNIT: SEEMS TO OCCUR AS DIMER, TETRAMERS, AND LARGE OLIGOMERS
 OF IDENTICAL CHAINS.
 CC -1- SUBCELLULAR LOCATION: THIS IS THE MAJOR MEMBRANE PROTEIN OF
 SPIROPLASMA.

CC -1- PPM: ACYLATED BY A PALMITIC ACID GROUP ON A THR.
 CC -1- SIMILARITY: 75% TO THE S.CITRI SPIRALIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: M59366; AAA26588.1; -
 DR PIR: A36149; A36149.
 KW Transmembrane; Palmitate; Lipoprotein.
 FT TRANSMEMBRANE: PROBABLE.
 FT DOMAIN 1 164 EXTRACELLULAR.
 FT TRANSMEM 165 184 POTENTIAL.
 FT DOMAIN 185 241 CYTOPLASMIC.
 FT LIPID 170 170 PALMITATE (POTENTIAL).
 SO SEQUENCE 241 AA; 25310 MM; 2865DFD CRC32;

SPIR_SPIME Length: 241 February 14, 2000 08:02 Type: P Check: 9527

1 KKLSTLAIV GVSAGTTSY VACNKTESNN LSKYKTIAP ATVAASTRKA

51 VTKPEIKTAL EAMVLKAVOG VKTATMADE QFEVYKNSKG TALETIDLEA
 101 GKVEYVLIQIT PAKDKTVIG ETRYIKVTLF KRGVETKVDI KDVTVEQTV
 151 GIASSTPKAV KDELNAVNT YATLAKAVLD ALQNTAPNAG ASDPEITNNG
 201 AEGDYENAKE VEYTYKAND SANISGOFK KAKYATAPT E

11AA_SEQUENCE 1.0 STANDARD; PRT; 134 AA.

ID SRP_CHLPS
 AC P28164: (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE SULFUR-RICH PROTEIN.
 DE SRP.
 OS Chlamydia psittaci.
 CC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-6BC;
 RX MEDLINE: 91267949.
 RA EVERETT K.D.E., HATCH T.P.;
 RT "Sequence analysis and lipid modification of the cysteine-rich
 RT envelope proteins of Chlamydia psittaci 6BC.";
 RL J. Bacteriol. 173:3821-3830(1991).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: M61116; AAB61620.1; -
 SO SEQUENCE 134 AA; 14752 MM; CC0422F5 CRC32;

SRP_CHLPS Length: 134 February 14, 2000 08:02 Type: P Check: 9648

1 MGENSNNSIG SDVTSILQPG LEQVODEGV QVSLINSVLG WCRVHINPI
 51 KTSKIVQSRRA FOITWVWVGI ILIAGLAUT FVLQGLQGN AFLFLPAVI
 101 GLVKLATSV CMERCTPER WRLCKRLMOQ LKIF

11AA_SEQUENCE 1.0 STANDARD; PRT; 352 AA.

ID SSAU_SALTY
 AC P60659: (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-DEC-1998 (Rel. 39, Last annotation update)
 DE SECRETION SYSTEM APPARATUS PROTEIN SSAU.
 GN SSAU.
 OS Salmonella typhimurium.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Salmonella.
 RN [1]

RC SEQUENCE FROM N.A.
 RP STRAIN-LT2;
 RX MEDLINE: 97285756.
 RA HENSEL M., SHEA J.E., RAUPACH B., MONACK D., FALKOW S., GLEESON C.,
 RA KUBO T., HOLDEN D.W.;
 RT "Functional analysis of ssau and the ssak/v operon, 13 genes encoding
 RT components of the type III secretion apparatus of Salmonella
 RT pathogenicity island 2.";
 RL Mol. Microbiol. 24:155-167(1997).
 CC -1- FUNCTION: PART OF A TYPE III SECRETION SYSTEM.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE FLMH/HRRP/YSCU/SPAS FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

DR PIR: A41795; A41795.
DR GCRDB; GCR_0272; -.
DR MIM: 182451; -.
DR PROSITE: PS00237; G-PROTEIN_RECEPTOR; 1.
DR PFAM: PF00001; 7cm_1; 1.
KW G-protein coupled receptor; Transmembrane; glycoprotein;
KW Multigene family; Lipoprotein; Palmitate.
FT DOMAIN 1 56
FT TRANSMEM 57 84 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 85 94 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 95 120 2 (POTENTIAL).
FT DOMAIN 121 131 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 132 153 3 (POTENTIAL).
FT DOMAIN 154 175 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 176 196 4 (POTENTIAL).
FT DOMAIN 197 219 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 220 244 5 (POTENTIAL).
FT DOMAIN 245 270 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 271 296 6 (POTENTIAL).
FT DOMAIN 297 303 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 304 327 7 (POTENTIAL).
FT DOMAIN 328 391 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 4 44 4
FT CARBOHYD 48 48 POTENTIAL.
FT DISULFID 130 208 BY SIMILARITY.
FT LIPID 339 339 PALMITATE (POTENTIAL).
SQ SEQUENCE 391 AA; 42686 MW; 28C01B27 CRC32;
SSRL_HUMAN Length: 391 February 14, 2000 08:02 Type: P Check: 8652
1 MFNGTASASP SSSPSPSGS CGEGGSGSRP GAGAADGME PGRNASONT
51 LSGQGSAIL ISFIYSYCL VGICGNSMT YVILRYAKMK TATNIYILNL
101 AIDELMLMS VPFLYSLIL RHMFECALLC RLVSVDVNV MFTSIYCLNV
151 LSVDRVAVV HPKARARR PIYAKVNVG VWVLSLVLIL PIYVESRTVA
201 NSGTCYACNM LMPEDACRM VGIVLYTFLM GFLPEYGAIC LCVLLIAMK
251 RMYALKAGMO ORKRSERKIT LMYMVAVYVF VICMPPFYV QLVNVAEED
301 DATVSQLSVI LGVANSCLNP ILYGELISDNF KRFSORILCL SWMDNAAEED
351 VDYVATALKS RAYSVEDFQ ENIESGQVRR NGTCSTRIT L
1.0
ID SSRL_MOUSE STANDARD: PRT; 391 AA.
AC P30873;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SOMATOSTATIN RECEPTOR TYPE 1 (SSIR) (SRIF-2).
DE SSTR1 OR SMSSTR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92108031.
RA YAMADA Y.; POST S.R.; WANG K.; TAGER H.S.; BELL G.I.; SEINO S.;
RT Cloning and functional characterization of a family of human and
RT mouse somatostatin receptors expressed in brain, gastrointestinal
RT tract, and kidney".
RT Proc. Natl. Acad. Sci. U S A. 89:251-255(1992).
RL -1- FUNCTION: RECEPTOR FOR SOMATOSTATIN WITH HIGHER AFFINITY FOR
CC SOMATOSTATIN 14 THAN 28. THIS RECEPTOR IS COUPLED VIA PERTUSSIS
CC TOXIN SENSITIVE G PROTEIN TO INHIBITION OF ADENYLATE CYCLASE. IN
CC ADDITION IT STIMULATES PHOSPHOTRANSFERASE AND NA+/H+
CC EXCHANGER VIA PERTUSSIS TOXIN INSENSITIVE G PROTEINS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: JEJUNUM AND STOMACH.

CC -----
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M81831; AAAS8255.1; -
CC PIR: C41795; C41795.
CC GCRDB: GCR_0273; -
CC MGI: MGI:98327; SMSYRL.
CC PROSITE: PS00237; G-PROTEIN_RECEPTOR. 1.
CC PFM: PFM:PF00001; 7tm_1; 1.
CC G-protein coupled receptor; Transmembrane; Glycoprotein;
CC Multigene family; Lipoprotein; Palmitate.
CC DOMAIN 1 56
CC TRANSMEM 57 84
CC TRANSMEM 85 94
CC TRANSMEM 95 120
CC TRANSMEM 121 131
CC TRANSMEM 132 153
CC TRANSMEM 154 175
CC TRANSMEM 176 196
CC TRANSMEM 197 219
CC TRANSMEM 220 244
CC TRANSMEM 245 270
CC TRANSMEM 271 296
CC TRANSMEM 297 303
CC TRANSMEM 304 327
CC TRANSMEM 328 391
CC TRANSMEM 391 44
CC CARBOHYD 44 44
CC CARBOHYD 48 48
CC CARBOHYD 130 208
CC DISULFID 139 339
CC LIPID 339 339
CC BY SIMILARITY.
CC PALMITATE (POTENTIAL).
CC SEQUENCE 391 AA; 42718 MW; BD335205 CRC32;
SSRL_MOUSE Length: 391 February 14, 2000 08:02 Type: P Check: 8110 ..
1 MPPNGTASSP SSSPSPSGS CGEGACSRG GSGADGME PGRNASNGT
51 LSEGGGSAIT ISFTYSVCL VGLGNSVY YVILRYAKK TATNIYILNL
101 AIADELIMS VPELVSTLL RHMPGALLC RLVSVDVAVN MFTSIYCLIV
151 LSVDRYAVV HPKARYR PTVAKVNLG VWVLSLVIL PIVESRTAA
201 NSDGTACNM IMPEPAQRL VGFVLYTFM GFLPVGAI C LVYLIIAKM
251 RMVALAGNO ORKRSERKIT LVMVMVYVF VICMPFYV QLVNVEAEO
301 DATVSQSLVY LGYANSCANP ILVGFSLDNF KRSFORICL SMDNAAEAP
351 VDYATATLKS RAYSVEDFOP ENLESGGVF NGTCSARIST L
11AA_SEQUENCE 1.0 STANDARD: PRT: 391 AA.
ID SSRL_RAT
AC P28646;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE SOMATOSTATIN RECEPTOR TYPE 1 (SSIR) (SRIF-2).
GN SSTRL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WISTAR; TISSUE-BRAIN;
RX MEDLINE: 92096119.

RA MEYERHOF W., PAUST H.J., SCHOENROCK C., RICHTER D.;
RT "Cloning of a cDNA encoding a novel putative G-protein-coupled
RT receptor expressed in specific rat brain regions.";
RT DNA Cell Biol. 10:689-694(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE: 93016064.
RA LI X.-J., FORTE M., NORTH R.A., ROSS C.A., SNYDER S.H.;
RT "Cloning and expression of a rat somatostatin receptor enriched in
RT brain.";
RT J. Biol. Chem. 267:21307-21312(1992).
CC -1- FUNCTION: RECEPTOR FOR SOMATOSTATIN WITH HIGHER AFFINITY FOR
CC SOMATOSTATIN-14 THAN -28. THIS RECEPTOR IS COUPLED TO
CC PHOSPHOTRANSFERASE PHOSPHATASE AND NA+/H+ EXCHANGER VIA PERTUSSIS
CC TOXIN INSENSITIVE G PROTEINS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: BRAIN, PITUITARY, ISLET, JEJUNUM, STOMACH,
CC HEART, SPLEEN.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M62314; CAA44193.1; -
CC EMBL: M97656; -; NOT_ANNOTATED_CDS.
CC PIR: A39297; A39297.
CC PIR: A43102; A43102.
CC GCRDB: GCR_0183; -
CC GCRDB: GCR_0628; -
CC PROSITE: PS00237; G-PROTEIN_RECEPTOR. 1.
CC PFM: PFM:PF00001; 7tm_1; 1.
CC G-protein coupled receptor; Transmembrane; Glycoprotein;
CC Multigene family; Lipoprotein; Palmitate.
CC DOMAIN 1 56
CC TRANSMEM 57 84
CC TRANSMEM 85 94
CC TRANSMEM 95 120
CC TRANSMEM 121 131
CC TRANSMEM 132 153
CC TRANSMEM 154 175
CC TRANSMEM 176 196
CC TRANSMEM 197 219
CC TRANSMEM 220 244
CC TRANSMEM 245 270
CC TRANSMEM 271 296
CC TRANSMEM 297 303
CC TRANSMEM 304 327
CC TRANSMEM 328 391
CC TRANSMEM 391 44
CC CARBOHYD 44 44
CC CARBOHYD 48 48
CC CARBOHYD 130 208
CC DISULFID 139 339
CC LIPID 339 339
CC BY SIMILARITY.
CC PALMITATE (POTENTIAL).
CC SEQUENCE 391 AA; 42746 MW; 8AF64505 CRC32;
SSRL_RAT Length: 391 February 14, 2000 08:02 Type: P Check: 9240 ..
1 MPPNGTASSP TSSPSSPGS CGEGVCSRG GSGADGME PGRNASNGT
51 LSEGGGSAIT ISFTYSVCL VGLGNSVY YVILRYAKK TATNIYILNL
101 AIADELIMS VPELVSTLL RHMPGALLC RLVSVDVAVN MFTSIYCLIV
151 LSVDRYAVV HPKARYR PTVAKVNLG VWVLSLVIL PIVESRTAA
201 NSDGTACNM IMPEPAQRL VGFVLYTFM GFLPVGAI C LVYLIIAKM


```

251 RYVALKAGMO ORKRSERKIT LMYVWYVWF VICWPEFYV QLVNVEAQD
301 DATVSQSLSVI LGVANSCLNP ILYGFLSDNF KRSFORILCL SMDNNAEPP
351 VDYVATPLKS RAYVEDFOP ENLESGVFR NGTCASRIST L

!!AA_SEQUENCE 1.0 STANDARD: PRT: 183 AA.
ID SSRB_CANPA
AC P23436;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE TRANSLOCOCN-ASSOCIATED PROTEIN, BETA SUBUNIT PRECURSOR (TRAP-BETA)
DE (SIGNAL SEQUENCE RECEPTOR BETA SUBUNIT) (SSR-BETA) (GP25H).
GN SSR2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Carnivora; Fissipedia; Canidae; Canis.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 18-50 AND 83-120.
RX MEDLINE; 91115924.
RA GOERLICH D., PREHN S., HARTMANN E., HERZ J., OTTO A., KRAFT R.,
RA WIEDMANN M., KNEPPEL S., DOBERSTEIN B., RAPPOPORT T.A.;
RT "The signal sequence receptor has a second subunit and is part of a
RT translocation complex in the endoplasmic reticulum as probed by
RT bifunctional reagents."
RL J. Cell Biol. 111:2283-2294(1990).
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 18-42; 59-106 AND 140-162.
RX MEDLINE; 92011761.
RA LODHA I., RINDRESS D., CAMERON P.H., CH W.-J., DOHERTY J.J., II,
RA LOVARD D., BELT A.W., DYKARD D., THOMAS D.I., BERGENON J.J.W.;
RT "SR alpha and associated calnexin are major calcium binding proteins
RT of the endoplasmic reticulum membrane."
RL J. Biol. Chem. 266:19599-19610(1991).
CC -1- FUNCTION: TRAP PROTEINS ARE PART OF A COMPLEX WHOSE FUNCTION IS TO
CC BIND CA(2+) TO THE ER MEMBRANE AND THEREBY REGULATE THE RETENTION
CC OF ER RESIDENT PROTEINS.
CC -1- SUBUNIT: HETEROTETRAMER OF TRAP-ALPHA, TRAP-BETA, TRAP-DELTA AND
CC TRAP-GAMMA.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ENDOPLASMIC
CC RETICULUM.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; X53529; CA37609.1; -.
DR EMBL; X53591; CA37661.1; -.
DR PIR; A36579; A36579.
DR PIR; B37273; B37273.
DR PIR; S15510; S15510.
KW Glycoprotein; Signal; Endoplasmic reticulum; Transmembrane.
FT SIGNAL 1 183
FT CHAIN 1 183
FT SUBUNIT.
FT TRANSLOCOCN-ASSOCIATED PROTEIN, BETA
FT SUBUNIT.
FT LUMENAL (POTENTIAL).
FT TRANSMEM 150 169
FT DOMAIN 170 183 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 88 88
FT CARBOHYD 104 104 POTENTIAL.
SQ SEQUENCE 183 AA; 20100 MW; 977848EB CRC32;

SSRB_CANPA Length: 183 February 14, 2000 08:02 Type: P Check: 8392

1 MRLASVYLA LFAVSHAEBG ARLLAKSKSL NRYAVEGDL TLQYNIYNG
51 SSAIDVVELS DDSFPEDFG IVSGMLNVKM DRIPASVNS HTVYLRPLKA

101 GYFNFTSATV TYLAOEDGPV VIGFTSAPGO GGLAOREFD RRPSPHLDW
151 AAFGVMTPLS IGIPLLMYS SKRKYDTPKS KKN

!!AA_SEQUENCE 1.0 STANDARD: PRT: 183 AA.
ID SSRB_HUMAN
AC P43308;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE TRANSLOCOCN-ASSOCIATED PROTEIN, BETA SUBUNIT PRECURSOR (TRAP-BETA)
DE (SIGNAL SEQUENCE RECEPTOR BETA SUBUNIT) (SSR-BETA).
GN SSR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-COLON CARCINOMA;
RX MEDLINE; 94114564.
RA BODESCOT M., BRISON O.;
RT "Cloning and sequence analysis of the beta subunit of the human
RT translococn-associated protein."
RL Biochim. Biophys. Acta 1217:101-102(1994).
CC -1- FUNCTION: TRAP PROTEINS ARE PART OF A COMPLEX WHOSE FUNCTION IS TO
CC BIND CA(2+) TO THE ER MEMBRANE AND THEREBY REGULATE THE RETENTION
CC OF ER RESIDENT PROTEINS.
CC -1- SUBUNIT: HETEROTETRAMER OF TRAP-ALPHA, TRAP-BETA, TRAP-DELTA AND
CC TRAP-GAMMA.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ENDOPLASMIC
CC RETICULUM.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; X74104; CA52207.1; -.
DR MIM; 600867; -.
KW Glycoprotein; Signal; Endoplasmic reticulum; Transmembrane.
FT SIGNAL 1 17
FT CHAIN 1 183
FT SUBUNIT.
FT TRANSLOCOCN-ASSOCIATED PROTEIN, BETA
FT SUBUNIT.
FT LUMENAL (POTENTIAL).
FT TRANSMEM 150 169
FT DOMAIN 170 183 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 88 88
FT CARBOHYD 104 104 POTENTIAL.
SQ SEQUENCE 183 AA; 20135 MW; B2E82B37 CRC32;

SSRB_HUMAN Length: 183 February 14, 2000 08:02 Type: P Check: 8093

1 MRLSEVYLA LFAVQAEBG ARLLAKSKSL NRYAVEGDL TLQYNIYNG
51 SSAIDVVELS DDSFPEDFG IVSGMLNVKM DRIPASVNS HTVYLRPLKA
101 GYFNFTSATV TYLAOEDGPV VIGFTSAPGO GGLAOREFD RRPSPHLDW
151 AAFGVMTPLS IGIPLLMYS SKRKYDTPKT KKN

!!AA_SEQUENCE 1.0 STANDARD: PRT: 214 AA.
ID STL4_SOLTU
AC Q41495;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE STL14 PROTEIN PRECURSOR.
GN STL14.
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

CC euphylliphytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 CC core eudicots; Asteridae; easterids I; Solanales; Solanaceae;
 CC Solanum.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV, DAPIRA; TISSUE-PISTIL;
 RX MEDLINE: 96197407
 RA VAN ELDIK G.J., WIGLEMS M., RUTTER R.K., VAN HERPEN M.M.A.,
 RA SCHRUMEN J.A.H., WIGLEMS G.J.;
 RT "Molecular analysis of a pistil-specific gene expressed in the stigma
 and cortex of Solanum tuberosum."
 RL Plant Mol. Biol. 30:171-176(1996).
 CC -1- FUNCTION: MAY PROTECT THE OUTER TISSUES OF THE PISTIL FROM
 CC -1- PATHOGEN ATTACK.
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE STIGMA AND STYLAR
 CC CORTEX THROUGHOUT PISTIL DEVELOPMENT. NOT EXPRESSED IN OTHER
 CC ORGANS.
 CC -1- DEVELOPMENTAL STAGE: ACCUMULATES IN THE PISTIL AROUND 120 HOURS
 CC BEFORE ANTHESIS AND INCREASES TOWARDS THE END OF FLOWER
 CC DEVELOPMENT, WITH A MAXIMUM AT ANTHESIS.
 CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
 CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X82652; CA57976.1; -
 DR HSRP: P04284; ICFE
 DR PROSITE: P50109; SCP_AG5_PRL_SC7_1; FALSE_NEG.
 DR PROSITE: P50109; SCP_AG5_PRL_SC7_2; FALSE_NEG.
 DR PIRAF: P00188; SCP; 1.
 KM Repeat: Signal.
 FT CHAIN 1 19 POTENTIAL.
 FT REPEAT 20 214 STS14 PROTEIN.
 FT DOMAIN 13 19 3.5 X 2 AA TANDEM REPEAT OF Y-I.
 FT DOMAIN 59 65 POLY-PRO.
 SQ SEQUENCE 214 AA: 23887 MW: 06BC0717 CRC32:
 SFL4_SQ_LU Length: 214 February 14, 2000 08:02 Type: P Check: 5285 ..
 1 MEVLSTAMAC LVYIYIYD EERKRELVK NMTLLEFQ FLITLASSL
 51 THISQIVPP PPPPTSNAT PPSRAQEF LDAHNRSEV GVGPLTSPM
 101 LAKETSLVR YORDKONCF ANLSNGKYG NQWASGVV TPRMVDWV
 151 AKKRYENEN NSCTGDDCKG VYQIYWKKS IELGAGNCTC YEPALTYC
 201 FYNPCNVIG EKP
 11AA_SEQUENCE 1.0
 ID SUFI_HAEN STANDARD: PRT: 311 AA.
 AC P44847
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE SUFI PROTEIN HOMOLOG PRECURSOR.
 GN SUFI OR H10733.
 OS Haemophilus influenzae.
 CC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 CC Haemophilus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RD / KW20;
 RX MEDLINE: 95350630.
 RA FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
 RA KERAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
 RA MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCARNE J.D.,

RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODER A., KELLEY J.M.,
 RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,
 RA UTERBACK T.R., HANNA M.C., NGUYEN D.T., SAIDER D.M., BRANDON R.C.,
 RA FINE L.D., FITCHMAN J.L., EHRMAN J.L., GEORGEAN N.S.M.,
 RA GNEHM C.L., McDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
 RA WENTER J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus
 influenzae Rd."
 RL Science 269:496-512(1995).
 CC -1- SUBCELLULAR LOCATION: PERIPLASMIC (POTENTIAL).
 CC -1- SIMILARITY: STRONG, TO E.COLI AND S.TYPHIMURIUM SUFI.
 CC -1- SIMILARITY: CONTAINS ONE PLASTOCYANIN-LIKE DOMAIN.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U32756; AAC2390.1; -
 DR TIGR: H10733; -
 DR PIRAF: P00188; SCP; 1.
 KM Repeat: Signal.
 FT CHAIN 1 28 BY SIMILARITY.
 FT SIGNAL 29 311 SUFI PROTEIN HOMOLOG.
 FT DOMAIN 68 164 PLASTOCYANIN-LIKE.
 SQ SEQUENCE 311 AA: 34496 MW: 8065147D CRC32:
 SUFI_HAEN Length: 311 February 14, 2000 08:02 Type: P Check: 4 ..
 1 MRLSRQLL KTAISTALS TVAPPLAAS REKLVPPLI EVRGRIPLV
 51 TMQETNPPLD GSHNTVWGF NGNYLGPTIK IKSSEFALIN YHNNIDPSVA
 101 LSIQGLASG ELFGAARVL KKGESNAPV PIEOPASCV YRSATLANSA
 151 YQYRGLAGM WLEDEOSLK ANLEPNKGYD DIPLIQDM EFNQGLOLF
 201 QNPFHFNLR LVNGIEAPY LDVARGWRL RLINASTARA YDLRLDNDQ
 251 MLIAQDLGF LPAKSVKSL VLSGERAEI LVNMKLTLY LSLAEVAVAC
 301 TKNKYYVLR R
 11AA_SEQUENCE 1.0
 ID SUH3_RAT STANDARD: PRT: 285 AA.
 AC P50235;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE ALCOHOL SULFOTRANSFERASE (ST) (ST-60).
 DE SULFOTRANSFERASE (ST) (ST-60).
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
 CC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE-LIVER;
 RX MEDLINE: 94306585.
 RA WATABE T., OGURA K., SATSUBARA M., OKUDA H., HIRATSUKA A.;
 RT "Molecular cloning and functions of rat liver hydroxysteroid
 Sulfotransferases catalysing covalent binding of carcinogenic
 RT polycyclic arylmethanols to DNA."
 RL Chem. Biol. Interact. 92:87-105(1994).
 CC -1- FUNCTION: CATALYSES SULFATION OF HYDROXYSTEROIDS AND XENOBIOTICS
 CC (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: 3'-PHOSPHOADENYLYLSULFATE + AN ALCOHOL -
 CC ADENOSINE 3',5'-BISPHOSPHATE + AN ARYL SULFATE.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- INDUCTION: INDUCED BY ESTROGENS AND SUPPRESSED BY ANDROGENS.
 CC EXPRESSION IS UNDER THE INFLUENCE OF PITUITARY GROWTH HORMONE AND
 CC THYROID HORMONE.

CC -1- SIMILARITY: BELONGS TO THE SULFOTRANSFERASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: D14989; BAA03634.1; -
 CC DR PFM: PFO0685; Sulfotransferase; 1.
 CC DR Transferase; Steroid metabolism.
 CC FT BINDING 249 255 PAPS BINDING SITE (POTENTIAL).
 CC SEQUENCE 285 AA; 33531 MW; DF4FC185 CRC32;
 SO
 SUB3_RAT Length: 285 February 14, 2000 08:02 Type: P Check: 800 ..
 1 MMSDVTWEG IPPPAFWFSK EILENSCKKF VKEDDILIL TYPKSGTNL
 51 IIVVLIQTK GDPKWIOSMP IMDRSPWET GSGYDKLTKM EGPRLMTSHL
 101 PHLFSKSLF SSRKAVIYLI RNPROYLSA YFFWSKALE KRPDSIGTY
 151 EWFLLGNVAY GSWFEHIRGW LSNRENDNL VLYEDMKD TMSIKRID
 201 FLGKLEPDE LNLVLYKTSF QVAKENNSN YLMEKELIL TGFTNRKGT
 251 TDMKRNHTV AQAEAFDKEV QEKMAFPFG MFPWE
 11AA_SEQUENCE 1.0
 ID T2C1_CHV13 STANDARD; PRT; 278 AA.
 AC P52283;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE TYPE II RESTRICTION ENZYME CVIJI (EC 3.1.21.4) (ENDONUCLEASE CVIJI)
 DE (R.CVIJI).
 OS Chlorella virus IL-3A (CV-IL3A).
 OS Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Phycodnavirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 95331597.
 RA SKOMRON P.M., SWAMINATHAN N., MCMASTER K., GEORGE D., VAN ETEN J.L.,
 RA MEAD D.A.;
 RT "Cloning and applications of the two/three-base restriction
 RT endonuclease R.CviJI from IL-3A virus-infected Chlorella";
 RL Gene 137:37-41(1995).
 RN [2]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-15.
 RX MEDLINE: 96252246.
 RA SWAMINATHAN N., MEAD D.A., MCMASTER K., GEORGE D., VAN ETEN J.L.,
 RA SKOMRON P.M.;
 RT "Molecular cloning of the three base restriction endonuclease R.CviJI
 RT from eukaryotic Chlorella virus IL-3A";
 RL Nucleic Acids Res. 24:2463-2469(1996).
 CC -1- CATALYTIC ACTIVITY: RECOGNIZES THE DOUBLE-STRANDED SEQUENCE RGCY
 CC AND CLEAVES AFTER G-2. IN THE PRESENCE OF ATP, THERE IS A
 CC RELAXATION OF ITS SPECIFICITY AND IT CAN CLEAVE RCGN AND YGCY, BUT
 CC NOT YGCR.
 CC -1- COFACTOR: REQUIRES MAGNESIUM.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U09001; AAC55064.1; -
 CC DR REBASE: RB00873; CviJI
 CC DR Hydrolyase; Endonuclease; Nuclease; Restriction system; Magnesium.
 KW

SO SEQUENCE 278 AA; 31633 MW; D23193CE CRC32;
 T2C1_CHV13 Length: 278 February 14, 2000 08:02 Type: P Check: 7233 ..
 1 MEEKRIALI EKORLADEKI ASGRIKIRI SINATHEHE FVKYNSAFV
 51 GPATFVVDI KNNKREIHN VPRELOGS KAKSPAYD REYKPKADI
 101 AAVDITKDV AMISHKASEG YQOYLKISG NKFTGKELE EVLSFKRVV
 151 SMAPSKIMP ANKTVSPK SNLIKNOAIF GEDYCKPKR DNVDIGGR
 201 PIIRKRSIL YLTFGFSAL NGHLENFTGK HEPVYVTE RSSGRSITT
 251 VVNGVTKNL REFIHPYNEV SSKTORIM
 11AA_SEQUENCE 1.0
 ID TCF1_HUMAN STANDARD; PRT; 269 AA.
 AC P36402;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE T-CELL-SPECIFIC TRANSCRIPTION FACTOR 1 (TCF-1) (T-CELL FACTOR 1)
 DE (TRANSCRIPTION FACTOR-7).
 GN TCF7 OR TCF1.
 OS Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
 CC Eutheria; Primates; Catarrhini; Homnidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 91114695.
 RA VAN DE WETERING M.L., OOSTERWEGEL M.A., DOOIJES D., CLEVERS H.;
 RT "Identification and cloning of TCF-1, a T lymphocyte-specific
 RT transcription factor containing a sequence-specific HMG box";
 RL EMBO J. 10:123-132(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 92235082.
 RA VAN DE WETERING M.L., OOSTERWEGEL M.A., HOLSTEGE F., DOOIJES D.,
 RA SUJCKERBUIJK R., GEURTS VAN KESSEL A., CLEVERS H.;
 RT "The human T cell transcription factor-1 gene. Structure,
 RT localization, and promoter characterization";
 RL J. Biol. Chem. 267:8530-8536(1992).
 CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR INVOLVED IN T-CELL LYMPHOCYTE
 CC DIFFERENTIATION. IT BINDS TO THE T-LYMPHOCYTE-SPECIFIC ENHANCER
 CC ELEMENT (5'-WCAAG-3') FOUND IN THE PROMOTER OF THE CD3-EPSILON
 CC GENE.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- ALTERNATIVE PRODUCTS: THREE DIFFERENT FORMS THAT DIFFER IN THE
 CC C-TERMINAL SECTION ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: T-CELL SPECIFIC.
 CC -1- SIMILARITY: CONTAINS 1 HMG BOX.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X59869; CAA42526.1; -
 CC DR EMBL: X59870; CAA42527.1; -
 CC DR EMBL: X59871; CAA42528.1; -
 CC DR EMBL: X63901; CAB56795.1; -
 CC DR PIR: S16645; S16645.
 CC DR PIR: A38095; A38095.
 CC DR TRANSFAC: T00999; -
 CC DR TRANSFAC: T01001; -
 CC DR MIM: 189908; -
 CC DR PFM: PFO0505; HMG-box; 1.
 CC DR DNA-binding; Nuclear protein; Transcription regulation; Activator;
 KW Alternative splicing.

Mon Feb 14 08:07:43 2000

sp.cat

Page 124

FT	DNA_BIND	154	222	HNG_BOX.
FT	VARSBLIC	244	269	ENMNPREDKDNQGESLSMSSSSPA -> GGRNRAFGTYP
FT	VARSBLIC	244	269	EKAAPAFPELPVYL (IN ISOFORM B).
FT	VARSBLIC	244	269	ENMNPREDKDNQGESLSMSSSSPA -> DGSPKRCRRAR
FT	VARSBLIC	244	269	FELNQTQWCAQPC (IN ISOFORM C).
FT	SEQUENCE	269 AA:	30264 MM:	7B387D9 CRC32.
TCF1_HUMAN	Length:	269	February 14, 2000 08:02	Type: P Check: 5007 ..

1 MYEYVYSAF NLIMHPPPS GAGHPQPCP PLHKANOPH GYVOLSYEYH
51 FNSPHPTAP ADISQOYVR PLQETDLBSG YSLTSGSMQ LPHYSWETH
101 PSLMLSGVP GHAPALPBA IVPESGQEL QEPDRNLKTO AESKAKEXK
151 KPTIKRPLNA FMYLXKEMA KYINECTIKE SAILNQLGR RHWALSREED
201 AKYELARKE ROLHQOLYNG WSDARDNYCK KRSREKHQE STTEINWPRE
251 LKDGNGQESL SMSSSSPA

11:AA_SEQUENCE 1.0	
ID	TELT_HUMAN
AC	015373
STANDARD;	PRT; 167 AA

DD 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-DEC-1999 (Rel. 39, Last annotation update)
 DE TELETHONIN (TITIN CAP PROTEIN).
 OS Homo sapiens (Human).
 OM Kaplayca; Melzosa.
 OC Eukaryota; Metazoa.
 OC Chordata; Craniata; Vertebrata.
 OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 NC (1)

RP SEQUENCE FROM N.A.
RC TISSUE-SKELETAL MUSCLE:
RD MEDULLINE, 98010471.
RE VALLE G., FAULKNER G., DE ANTONI A., PACHIONI B., PALLAVICINI A.
RF PANDOLFI D., TISO N., TOPPO S., TREVISAN S., LANFRANCHI G.;
RG "Telothonh", a novel serosmic protein of heart and skeletal
RH muscle." RT
RL FEBS Lett. 415:163-168(1997).

RP SEQUENCE FROM N.A.
RA PAULAVICINI A., VALLE G., IANFRANCHI G.;
RT "Human telomeron genomic sequence.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBD databases
173

RP SEQUENCE FROM N.A.
RP MOES A., GAUTBL M.;
RT "Structure of the human telethonin gene."; [PubMed](#)
RT Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases
RL [4]
RN INTERACTION WITH TITIN.
RP

RA MUES A., VAN DER VEN P.F., YOUNG P., FURST D.O., GAUTEL M.;
RT "Two immunoglobulin-like domains of the z-disc portion of titin
TT interact in a conformation-dependent way with telethonin.";
RL FEBS Lett. 488:11-114(1998).
CC
CC -I- FUNCTION: MUSCLE ASSEMBLY REGULATING FACTOR.
CC -I- SUBCELLULAR LOCATION: SARCOMERIC.

 1. TISSUE SPECIFICITY: HEART AND SKELETAL MUSCLE.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR	EMBL; AJ000491; -
DR	EMBL; AJ010063; -
DR	EMBL; AJ011098; CAA09479.1; -

```

SO SEQUENCE 167 AA; 19052 MW; EBC8B76C CRC32;
TEL_HUMAN Length: 167 February 14, 2000 08:02 Type: P Check: 3375
1 MATSELSCEV SEENCERREA FAEAKKDLTL STRPEGCSI HEEDQRET
51 YHQGGCCQYL VQSPMLAMR KQILRGQIQE YQPIYQVLD LPITPAKMG
101 AIKEKEEDTP IQQLQELALE TALGGCCVDR QEVAETIKOL PPVVPVSKPG
151 ALRRSLRSM SQEAQRG

```

1 MATSLSJSEV SEENCERREA FMAEMKDLTL STREPGCSL HEEDTQHEH
51 YHQOQOCQVLR VORSPLMNR MGLGRGJQE YQLPQVRVLP LPFTPRXMKM
101 AKKEREDPT IQOQELLALF TALGQCQVDR QZVAETIKOL PPVVPVSKPCQ
151 ALRRLSLRSM SOZAOQNG

11AA_SEQUENCE 1.0	
ID	STANDARD;
TELT_MOUSE	PRT;
167 AA.	

15-DEC-1999 (Rel. 39, Created)
 15-DEC-1999 (Rel. 39, Last sequence update)
 15-DEC-1999 (Rel. 39, Last annotation update)
 TELEPHONIN (TITIN Cap PROTEIN).
 Mus musculus (Mouse).
 Mus musculus (Mouse).
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 11

RP SEQUENCE FROM N.A.
RC TISSUE-DIAPHRAGM;
RA IEOVELELLA C.; FORMENTIN E.; VALLE G.; LANFRANCHI G.;
RT "Skeletal muscle transcripts characterization in Homo sapiens and Mus
R1 musculi";
RL Submitted (FEB-1998) to the EMBL/Genbank/DBJ databases.
[2]

RP SEQUENCE FROM N.A.
RC TISSUE=HEART;
RA KOLMERER B. ;
RT "The titin cap protein - a novel protein essential for sarcomere
RT formation.";
RT Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC 1. FUNCTION: MUSCLE ASSEMBLY REGULATING FACTOR.
CC

```

CC      This swiss-prot entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC      at the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb.sib.ch).
CC
CC      -1 SUBUNIT: INTERACTS WITH ITIN.
CC

```

DR EMBL; AJ223854; CAA11585.1; -.
DR EMBL; Y15845; CAB38077.1; -.
SQ SEQUENCE 167 AA; 19078 MW; 06F264F1 CRC32

TELT_MOUSE Length: 167 February 14, 2000 08:02 Type: P Check: 5613

1 MATSELSCOV SEBNERERRA FMAEMKUTL SRPREGCSL HEDDITORHEE
51 YHRGGOCCOAV VARSPLVYVR LGILGRLOE YOLPYQVRLV LPITPLPKVAG
101 ASKEEBETP IOLRELLALE TALGGOCVER QDAVEITKOL PYYVPVSKPGC
151 PLRUTLSRSM SQEADRG

```
111AA_SEQUENCE 1.0
ID TRAM_AGR6 STANDARD; PRT; 102 AA.
```

01-NOV-1997 (Rel. 35, Created)
01-NOV-1997 (Rel. 35, Last sequence update)
01-NOV-1997 (Rel. 35, Last annotation update)
TRANSCRIPTIONAL REPRESSOR TRAM.

Agrobacterium tumefaciens.
plasmid pT1A6NC.
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group

OC Rhizobiaceae: Agrobacterium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 95173115.
 RA FLOU C., BURBEA M., WINANS S.C.;
 RT "Activity of the Agrobacterium T1 plasmid conjugal transfer regulator
 RT T1 is inhibited by the product of the *traM* gene";
 RT J. Bacteriol. 177:1367-1373(1995).
 CC -1- FUNCTION: NEGATIVELY REGULATES CONJUGATION AND *TRA* GENES
 CC EITHER BIND OR MODIFY *TRAR* OR *RAI* MAKING THEM UNAVAILABLE
 CC ALTERATIVELY, *TRAM* MAY BIND *TRA* PROMOTERS PREVENTING *TRAR*
 CC ACTIVATION.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U43674; AAC28120.1; -
 CC EMBL: U16786; AAA64838.1; -
 CC Transcription regulation; Repressor; Plasmid.
 KW SEQUENCE 102 AA; 11401 MW; AAC3EDD6 CRC32;
 SO
 TRAM_AGR6 Length: 102 February 14, 2000 08:02 Type: P Check: 4597 ..
 1 MELEBANATK KVELRPLIGL TRGLPTDLE TITIDAIRTH RLVEKADL
 51 FOALRETYKT GQACGGPOHI RYLEASIEHM AQMSALNTLY SILGFIPKYV
 101 VN
 !!AA_SEQUENCE 1.0
 ID TRM_SYNY3 STANDARD: PRT: 231 AA.
 AC P72828;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT TRNA (GUANINE-N1)-METHYLTRANSFERASE (EC 2.1.1.31) (MIG-
 DE METHYLTRANSFERASE) (TRNA [GM37] METHYLTRANSFERASE).
 GN TRM OR SL1198.
 OS Synecocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 97061201.
 RA KANeko T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
 RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
 RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K.,
 RA OKUMURA S., SHIMO S., TAKEUCHI C., WADA T., WATANABE A.,
 RA YAMADA M., YASUDA M., TABATA S.;
 RT Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synecocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.;
 RL DNA Res. 3:109-136(1996).
 CC -1- FUNCTION: SPECIFICALLY METHYLATES GUANOSINE-37 IN VARIOUS TRNAS
 CC (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + TRNA = S-ADENOSYL-
 CC L-HOMOCYSTEINE + TRNA CONTAINING N(1)-METHYLGUANINE.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE RNA METHYLTRANSFERASE TRM FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Z98977; CAB11659.1; -
 CC DR Transferase; Methyltransferase; trna processing.
 KW SEQUENCE 415 AA; 47626 MW; 36B7C485 CRC32;
 SO
 TRM_SCHRO Length: 415 February 14, 2000 08:02 Type: P Check: 7997 ..
 1 MVSFLQKQ IIEGSKAFOP HSTRLOWPKS QDKVFVMSG GVDSSFSAYL
 51 LKQSYNVEG VPKRNWIDED SAPSCGPAER DMATVQVCK KINISRRRN
 101 FKEKWNIVF EPSIDLYENG LTPNDVSCN RQVFGALFD ALKKHENVV
 151 KQDWMLASGH YAKSVNIET NESHCIPFD KRKQITFLC TIRKELKNT
 201 IEPFLHWKE NVKKQASSAG FEELKQDS QGLCFVSPNV GKKRFKFLOR
 251 YLNFSDRIK VIAGKNVGE FSGNHGWSL TVSGRCGLSL POAOSFEYGR
 301 WYVKKDKIN NALYICRGT NELNSKICY LKDWKCGTK LONLEKSALS
 351 CEVVRHOOP LOPAKYWRN PESVRIHFD KORAVTPGOV IAVYVDCVL
 401 GGGVDTVEP EKDFD
 !!AA_SEQUENCE 1.0
 ID TRM_YEAST STANDARD: PRT: 417 AA.

DR EMBL: D90901; BAA16843.1; -
 KW Transferase; Methyltransferase; trna processing.
 SO SEQUENCE 231 AA; 25949 MW; B0CF12B1 CRC32;
 TRM_SYNY3 Length: 231 February 14, 2000 08:02 Type: P Check: 6210 ..
 1 MQFVLTLFP DEFTSPLOSG LLGKALEKAI ASYNILNPRD FTTKRRRD
 51 DEFGGVGM VIKPEPIFAA VESLPVLSKR EYLMTPOG PMDOLFREL
 101 TNYQVLIC GHYGVDEHV COLVREYSL GDEVLCGEI PALTLINGVI
 151 RLLEGVGKE ASLNASFST DLUDYPHTR PPFVRLAVP PYLLSGNHQA
 201 IAWRLRQOE ERTOGRRPDL WQKWDRPS P
 !!AA_SEQUENCE 1.0
 ID TRM_SCHPO STANDARD: PRT: 415 AA.
 AC O13947;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE PROBABLE TRNA (5-METHYLAMINOMETHYL-2-THIOURIDYLATE)-METHYLTRANSFERASE
 DE (EC 2.1.1.61).
 GN SPAC23H4.04.
 OS Schizosaccharomyces pombe (Fission yeast).
 OS Eukaryota; Fungi; Ascomycota; Archiascomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972;
 RA BROWN D., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WOOD V.;
 RU Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + TRNA = S-ADENOSYL-L-
 CC HOMOCYSTEINE + TRNA CONTAINING 5-METHYLAMINOMETHYL-2-
 CC THIOURIDYLATE.
 CC -1- SIMILARITY: BELONGS TO THE TRM FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Z98977; CAB11659.1; -
 CC DR Transferase; Methyltransferase; trna processing.
 KW SEQUENCE 415 AA; 47626 MW; 36B7C485 CRC32;
 SO
 TRM_SCHRO Length: 415 February 14, 2000 08:02 Type: P Check: 7997 ..
 1 MVSFLQKQ IIEGSKAFOP HSTRLOWPKS QDKVFVMSG GVDSSFSAYL
 51 LKQSYNVEG VPKRNWIDED SAPSCGPAER DMATVQVCK KINISRRRN
 101 FKEKWNIVF EPSIDLYENG LTPNDVSCN RQVFGALFD ALKKHENVV
 151 KQDWMLASGH YAKSVNIET NESHCIPFD KRKQITFLC TIRKELKNT
 201 IEPFLHWKE NVKKQASSAG FEELKQDS QGLCFVSPNV GKKRFKFLOR
 251 YLNFSDRIK VIAGKNVGE FSGNHGWSL TVSGRCGLSL POAOSFEYGR
 301 WYVKKDKIN NALYICRGT NELNSKICY LKDWKCGTK LONLEKSALS
 351 CEVVRHOOP LOPAKYWRN PESVRIHFD KORAVTPGOV IAVYVDCVL
 401 GGGVDTVEP EKDFD
 !!AA_SEQUENCE 1.0
 ID TRM_YEAST STANDARD: PRT: 417 AA.

AC 012093;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE PROBABLE TRNA (5-METHYLAMINOMETHYL-2-THIOURIDYLATE)-METHYLTRANSFERASE
 (EC 2.1.1.61).
 GN YD1033C OR D2761.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
 CC Saccharomycetaceae; Saccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=288C;
 RA PAULIN L., SAREN A.M.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + TRNA - S-ADENOSYL-L-
 CC HOMOCYSTEINE + TRNA CONTAINING 5-METHYLAMINOMETHYL-2-
 CC THIOURIDYLATE.
 CC -1- SIMILARITY: BELONGS TO THE TRMU FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z71781; CAA96456.1; -
 DR EMBL: Z74081; CAA98591.1; -
 RW Transferase: Methyltransferase; trna processing.
 SQ SEQUENCE 417 AA; 47049 MW; 4690E911 CRC32;
 TRMU_YEAST Length: 417 February 14, 2000 08:02 Type: P Check: 8415

1 MLARYNLIG RRSASPRPQ RLPKAFDNVI VAMSSGVDS VAAALPAGEF
 51 PNTGVTWQWQ WSESQSLDD GKEPCYERDW RDVNRVAKHL NTRVDKNEE
 101 ODYIVDFEP MARGYSEGST PNPDIGCKNF VKFGKLEWML DEKYGTGNWY
 151 LVTHYAVRW QEMNGKGLFH ILRSYRPRD QSYLLSQINS TVLSSILLPI
 201 GHLTKPEVRD LAKYAGLPTA EKPDQSGICF VNNSOHGKFK NFLKHYLPSS
 251 PGDIITVDPQ SGAKTWGRH DGLMSYTIQK KVGISMPQAD PNYQGTWYVS
 301 EKLDTWEIL IVRGRNPAL YSDTWRIENF SSLGPREDTI NAFQNTGALT
 351 LQFSLQVPP QIKSKLNRS ADNLDITLH ASKORATTPG QSCCLYIDDR
 401 VLGGPISHV NNNDTHA

11AA_SEQUENCE 1.0 STANDARD; PRT; 268 AA.
 ID TRNH_DATST
 AC P50165;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DE TROPIONONE REDUCTASE HOMOLOG (EC 1.1.1.-) (P23X)
 OS Datura stramonium (Jimsonweed) (Common thornapple).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 CC core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;
 CC Datura.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=ROOT;
 RC MEDLINE; 94022421.
 RA NAKAJIMA K., HASHIMOTO T., YAMADA Y.;
 RT "Two tropinone reductases with different stereospecificities are
 RT short-chain dehydrogenases evolved from a common ancestor";
 Proc. Natl. Acad. Sci. U.S.A. 90:9591-9595(1993).

CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 CC FAMILY (SDR).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L20475; AAA33280.1; -
 DR HSSP; P19992; 2HSD.
 DR PROSITE; P500061; ADH_SHORT; 1.
 DR PFAM; PF00106; adh_short; 1.
 DR PFAM; PF00678; adh_short_C2; 1.
 RW Oxidoreductase; NADP.
 FT ACT_SITE 167 167 NADP (BY SIMILARITY).
 FT ACT_SITE 21 45 BY SIMILARITY.
 SQ SEQUENCE 268 AA; 28588 MW; 0BB5A16C CRC32;
 TRNH_DATST Length: 268 February 14, 2000 08:02 Type: P Check: 5293

1 MAGREIIGGD RRMSLRGMTA IYTGTRGIG VAIVEELANF GAELYTCSSRS
 51 QNDLECLEX WRKKGFKVSG PQCVDSSISG RQTLMEVTS SPNGKINILI
 101 NNAGTIIPKE ATNFTAEYS IINGTNEAS VYLQDLAPL LKASGNASIV
 151 FNSNAGVIA VELSITVAS KGAINVYKS LACEAKRSI RVNAVAPWIT
 201 NTFILAAQ VPSQKNIES LIGRAAPKRA GEPSEVSIV TYLCPLASTY
 251 ITGQIICVDG GYTVNGFI

11AA_SEQUENCE 1.0 STANDARD; PRT; 312 AA.
 ID TRPE_CRYNE
 AC P27710;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DE 01-JUN-1994 (Rel. 29, Last annotation update)
 DE N-(5'-PHOSPHORIBOSYL)ANTHRANILATE ISOMERASE (EC 5.3.1.24) (PRAI).
 GN TRP1.
 OS Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Tremellales;
 CC Tremellaceae; Filobasidiella.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE; 93083987.
 RA PERFECT J.R., RUDE T.H., PENNING L.M., JOHNSTON S.A.;
 RT "Cloning the cryptococcus neoformans TRP1 gene by complementation in
 RT Saccharomyces cerevisiae";
 RL Gene 122:213-217(1992).
 CC -1- CATALYTIC ACTIVITY: N-(5'-PHOSPHO-BETA-D-RIBOSYL)-ANTHRANILATE -
 CC 1-(2-CARBOXYPHENYLAMINO)-1-DEOXY-D-RIBULOSE 5-PHOSPHATE.
 CC -1- PATHWAY: THIRD STEP IN TRIPHOHAN BIOSYNTHESIS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M74901; AAA51445.1; -
 DR PIR; JN0451; JN0451.
 DR HSSP; O56320; INSG.
 DR PFAM; PF00218; IGPS; 1.
 DR PFAM; PF00697; PRAI; 1.
 KW isomerase; tryptophan biosynthesis.
 SQ SEQUENCE 312 AA; 33070 MW; CSA4F8A8 CRC32;

```

TTF1_CRYME Length: 312 February 14, 2000 08:02 Type: P Check: 2342
1 MSTRSRVNA LNRDVLVCA LSGISHEV EKYKEGVG VLVGELMBA
51 SPTKFLSL IGLPPLVYS KRPPLVYKIG IRSTNDAKLA INNGALLGV
101 ILVPGTKRCI STTAREISA LVOSARSQSS SKPLPSLSS PMTQSALL
151 SSRRRPLVG VFONQSLSDI LSAVDEIGLD LVQHGDEPO AMAKTPVY
201 VVFRVSPBG IVRGEIRRP GLNQAILLDA GGASGGGGG KAPWEHAKR
251 LIOSEVSE GHVPLPYLA GGLTPENVQ ALNRLVAFG VMMSAVGSK
301 REGRRRLR RS

!!AA_SEQUENCE 1.0
ID TTF1_CANFA STANDARD: PRT: 371 AA.
AC P43698;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE THYROID TRANSCRIPTION FACTOR 1 (THYROID NUCLEAR FACTOR 1) (TTF-1)
DE (HOMEOBOX PROTEIN NKX-2.1).
GN TTF1 OR TTF1
OS Eukarya: Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Carnivora; Fissipedia; Canidae; Canis.
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE: 96034516.
RA VAN RENTERGHEM P.H.G., DRENIER S., VASSAR G., CHRISTOPHE J.;
RT "Study of TTF-1 gene expression in dog thyrocytes in primary
RT culture.";
RL Mol. Cell. Endocrinol. 112:83-93(1995).
CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS AND ACTIVATES THE
CC PROMOTER OF THYROID SPECIFIC GENES SUCH AS THYROGLOBULIN,
CC THYROPEROXIDASE, AND THYROTROPIN RECEPTOR. CRUCIAL IN THE
CC MAINTENANCE OF THE THYROID DIFFERENTIATION PHENOTYPE. MAY PLAY A
CC ROLE IN LUNG DEVELOPMENT AND SURFACTANT HOMEOSTASIS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: THYROID, LUNG AND CNS.
CC -1- SIMILARITY: BELONGS TO THE NK-2 FAMILY OF HOMEOBOX PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X77910; GMA54868.1;
DR HSSP: P23441; 1FTT.
DR TRANSFAC: T02098;
DR PROSITE: PS00027; HOMEOBOX_1; 1.
DR PROSITE: PS50071; HOMEOBOX_2; 1.
DR PFAM: PF00046; homeobox; 1.
KW Transcription regulation; Activator; Homeobox; DNA-binding;
KW Nuclear protein.
FT DNA_BIND 161 220 HOMEOBOX.
FT DOMAIN 234 243 POLY-GLY.
FT DOMAIN 246 253 POLY-GLN.
FT DOMAIN 294 303 POLY-ALA.
SQ SEQUENCE 371 AA: 38539 MW: 5662CA2D CRC32:

TTF1_CANFA Length: 371 February 14, 2000 08:02 Type: P Check: 4988
1 MMSPKRTP FVSVDILSP L ESKYKVMG GGLGAPLVA YRQQAAPPA
51 AAMQAHVGH HGAATRAVHM TAAGVPOLSH SAVGVCNKN LGNNSLPY
101 QDTMNSASG PGWYGANDP RPFAISRFKG PASGNMGM GGLSLDYS
151 KMAFLPSAP RRRRVLFQO AQYELERF RQKYLSPAE RHLASMIHL
201 TPTQKIMFO NHRKMKRQA KDKAAQOOLQ QDSGGGGGG GAGCPQOQA
251 QQSPRRVAV PVLVRGKPC QAGAPAGAG SLGGHMOQA QQAQAQAQA
301 AAATSVGSG PGIGAHPGHQ PGSAGQSPDL AHHAASPAL QGVSSLSHL
351 NSSGSDYGTM SCSTLLXGRT W

!!AA_SEQUENCE 1.0
ID TTF1_HUMAN STANDARD: PRT: 371 AA.
AC P43699; 014955; 014954;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE THYROID TRANSCRIPTION FACTOR 1 (THYROID NUCLEAR FACTOR 1) (TTF-1)
DE (HOMEOBOX PROTEIN NKX-2.1).
GN TTF1 OR TTF1 OR NKX2A.
OS Homo sapiens (Human).
OC Eukarya: Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE: 95226464.
RA SAJARDI A., TASSI V., DE FILIPPIS V., CIVITAREALE D.;
RT "Cloning and sequence analysis of human thyroid transcription factor
RT 1.";
RL Biochim. Biophys. Acta 1261:307-310(1995).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE: 95229626.
RA IKEDA K., CLARK J.C., SHAW-WHITE J.R., STAHLMAN M.T., BOUTELL C.J.,
RA WHITSETT J.A.;
RT "Gene structure and expression of human thyroid transcription
RT factor-1 in respiratory epithelial cells.";
RL J. Biol. Chem. 270:8108-8114(1995).
RN (3)
RP SEQUENCE FROM N.A.
RX MEDLINE:
RA OGUCHI H., KIMURA S.;
RT Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
RN (4)
RP SEQUENCE FROM N.A.
RX MEDLINE:
RA HAYMAN H., LIU H., DELEMOS R., MINO P.;
RT Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN (5)
RP SEQUENCE FROM N.A.
RX MEDLINE:
RA HAYMAN H., LIU H., JONES C., DELEMOS R., MINO P.;
RT Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
RN (6)
RP SEQUENCE FROM N.A.
RX MEDLINE:
RA ENDO T., OHNO M., NAKAZATO M.;
RT Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
RN (7)
RP FUNCTION: TRANSCRIPTION FACTOR THAT BINDS AND ACTIVATES THE
RN PROMOTER OF THYROID SPECIFIC GENES SUCH AS THYROGLOBULIN,
RN THYROPEROXIDASE, AND THYROTROPIN RECEPTOR. CRUCIAL IN THE
RN MAINTENANCE OF THE THYROID DIFFERENTIATION PHENOTYPE. MAY PLAY A
RN ROLE IN LUNG DEVELOPMENT AND SURFACTANT HOMEOSTASIS.
RN -1- SUBCELLULAR LOCATION: NUCLEAR.
RN -1- TISSUE SPECIFICITY: THYROID AND LUNG.
RN -1- SIMILARITY: BELONGS TO THE NK-2 FAMILY OF HOMEOBOX PROTEINS.
RN -----
RN This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial

```


entitles requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: X82850; CAAS8053.1; -
 CC EMBL: U19816; AAC50125.1; -
 DR EMBL: U19756; AAA86099.1; -
 DR EMBL: U43203; AAA89066.1; ALT_INIT.
 DR EMBL: U33749; AAB52381.1; -
 DR EMBL: D50740; BAA23529.1; -
 DR EMBL: D50739; BAA23527.1; -
 DR HSSP: P23441; 1FTT.
 DR TRANSFAC: T00857; -
 DR MIM: 600635; -
 DR PROSITE: PS00027; HOMEBOX_1; 1.
 DR PROSITE: PS50071; HOMEBOX_2; 1.
 DR PFM: PF00046; homeobox; 1.
 KW Transcription regulation; Activator; Homeobox; DNA-binding;
 FT DNA_BIND 161 220 HOMEBOX.
 FT DOMAIN 234 243 POLY-GLY.
 FT DOMAIN 246 253 POLY-GLN.
 FT VARSPLIC 112 125 POLY-ALA.
 FT VARSPLIC 112 125 GWYGANDPREFPAL -> V (IN ISOFORM 2).
 FT CONFLICT 49 49 P -> H (IN REF. 6).
 FT CONFLICT 61 61 H -> P (IN REF. 6).
 FT CONFLICT 158 158 S -> T (IN REF. 6).
 FT CONFLICT 161 161 R -> G (IN REF. 6).
 FT CONFLICT 226 227 QQ -> HE (IN REF. 5).
 SQ SEQUENCE 371 AA; 38596 MW; FEA215E7 CRC32;

TF1_HUMAN Length: 371 February 14, 2000 08:02 Type: P Check: 5439

1 MSMSPKHTP FVSVDLSPL EESYKKGME GGLGAPLAA YRQGAAPP
 51 AAMQOHAVGH HGAVTAAYHM TAAGVPLSH SAVGVCNNGN LGNMSLEPPY
 101 QDTMRNSASG PGWYGANDP RFPALSRFG PASGNNMSGM GGLGSLDVS
 151 KNAAPLPSAP RRRKRVLSQ AQYELERRF KQKYLSP EHLASMTL
 201 TPQYKIMFO NHRKMKRQA KDKAAQOQL QDSGGGGGG GTCPCQQA
 251 AQQSPRRVA PVLYKDKRP QAGAPAPGA SLQGHAAQA QHQAQAQA
 301 AAATVSGG AGLGAPGHQ PGSAGSPDL AHHAAPAL QGVVSLSH
 351 NSSGSDYGT MSCSTLYGRT W

11AA_SEQUENCE 1.0 STANDARD; PRT; 372 AA.
 ID TF1_MOUSE
 AC P50220;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE THYROID-TRANSCRIPTION FACTOR 1 (THYROID NUCLEAR FACTOR 1) (TTF-1)
 DE (HOMEBOX PROTEIN NKX-2.1).
 GN TTF1 OR TTF1 OR NKX-2.1 OR NKX-2.1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-129/SV; TISSUE-LIVER;
 RX MEDLINE: 95226463
 RA OUCHT H., PAN Y.T., KIMURA S.;
 RT "The complete nucleotide sequence of the mouse thyroid-specific
 enhancer-binding protein (T/EBP) gene: extensive identity of the
 deduced amino acid sequence with the human protein.";
 RL Blochm. Biophys. Acta 1261:304-306(1995).
 CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS AND ACTIVATES THE
 PROMOTER OF THYROID SPECIFIC GENES SUCH AS THYROGLOBULIN,
 THYROPEROXIDASE, AND THYROTROPIN RECEPTOR. CRUCIAL IN THE
 MAINTENANCE OF THE THYROID DIFFERENTIATION PHENOTYPE. MAY PLAY A

ROLE IN LUNG DEVELOPMENT AND SURFACTANT HOMEOSTASIS.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- TISSUE SPECIFICITY: THYROID, LUNG AND BRAIN.
 CC -1- SIMILARITY: BELONGS TO THE NK-2 FAMILY OF HOMEBOX PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL: U19755; AAA86100.1; -
 CC HSSP: P23441; 1FTT.
 DR TRANSFAC: T00859; -
 DR MGD: MGI:108067; TTF1.
 DR PROSITE: PS00027; HOMEBOX_1; 1.
 DR PROSITE: PS50071; HOMEBOX_2; 1.
 DR PFM: PF00046; homeobox; 1.
 KW Transcription regulation; Activator; Homeobox; DNA-binding;
 FT DNA_BIND 161 220 HOMEBOX.
 FT DOMAIN 234 244 POLY-GLY.
 FT DOMAIN 247 254 POLY-GLN.
 FT DOMAIN 295 304 POLY-ALA.
 SQ SEQUENCE 372 AA; 38570 MW; E047C32A CRC32;

TF1_MOUSE Length: 372 February 14, 2000 08:02 Type: P Check: 7622

1 MSMSPKHTP FVSVDLSPL EESYKKGME GGLGAPLAA YRQGAAPP
 51 AAMQOHAVGH HGAVTAAYHM TAAGVPLSH SAVGVCNNGN LGNMSLEPPY
 101 QDTMRNSASG PGWYGANDP RFPALSRFG PASGNNMSGM GGLGSLDVS
 151 KNAAPLPSAP RRRKRVLSQ AQYELERRF KQKYLSP EHLASMTL
 201 TPQYKIMFO NHRKMKRQA KDKAAQOQL QDSGGGGGG GAGCPCQQA
 251 AQQSPRRVA PVLYKDKRP QAGAPAPGA ASLQSHAQA AQAQAQAQA
 301 AAATVSGG GAGLGAPGHQ QPSAGSPDL LAHHAAPAG LQGVVSLSH
 351 LNSSGSDYGA MSCSTLYGR TW

11AA_SEQUENCE 1.0 STANDARD; PRT; 372 AA.
 ID TF1_RAT
 AC P23441; O08630; 070121;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE THYROID-TRANSCRIPTION FACTOR 1 (THYROID NUCLEAR FACTOR 1) (TTF-1)
 DE (HOMEBOX PROTEIN NKX-2.1).
 GN TTF1 OR TTF1 OR NKX-2.1 OR NKX-2.1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE: 91006063.
 RA GAZZIT S., PRICE M., DE FELICE M., DAMANTE G., MATTEI M.-G.,
 RA DI LAURO R.;
 RT "Thyroid nuclear factor 1 (TTF-1) contains a homeodomain and displays
 a novel DNA binding specificity.";
 RL EMBO J. 9:3631-3639(1990).
 CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS AND ACTIVATES THE
 PROMOTER OF THYROID SPECIFIC GENES SUCH AS THYROGLOBULIN,
 THYROPEROXIDASE, AND THYROTROPIN RECEPTOR. CRUCIAL IN THE
 MAINTENANCE OF THE THYROID DIFFERENTIATION PHENOTYPE. MAY PLAY A

RL Biochem. Biophys. Res. Commun. 204:1358-1363(1994).

RN [3]

RP STRUCTURE BY NMR OF 161-227.

RA MEDLINE: 94109576.

RX VIGILINO P., FOCOLARI F., FORMISANO S., BORTOLOTTI N., DAMANTE G.,

RA VIGILINO P., FOCOLARI F., FORMISANO S., BORTOLOTTI N., DAMANTE G.,

RA DI LAURO R., ESPOSITO G.,

RT "Structural study of rat thyroid transcription factor 1 homeodomain

RT (TF1-HD) by nuclear magnetic resonance."

RL FEBS Lett. 336:397-402(1993).

RN [4]

RP STRUCTURE BY NMR OF 161-227.

RA MEDLINE: 97054597.

RA ESPOSITO G., FOCOLARI F., DAMANTE G., FORMISANO S., TELL G.,

RA ESPOSITO G., FOCOLARI F., DAMANTE G., FORMISANO S., TELL G.,

RT "Analysis of the evolution structure of the homeodomain of rat thyroid

RT transcription factor" by 1H-NMR spectroscopy and restrained

RT molecular mechanics."

RL Eur. J. Biochem. 241:101-113(1996).

CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS AND ACTIVATES THE

CC PROMOTER OF THYROID SPECIFIC GENES SUCH AS THYROGLOBULIN,

CC THYROPEROXIDASE, AND THYROTROPIN RECEPTOR. CRUCIAL IN THE

CC MAINTENANCE OF THE THYROID DIFFERENTIATION PHENOTYPE. MAY PLAY A

CC ROLE IN LUNG DEVELOPMENT AND SURFACTANT HOMEOSTASIS.

CC -1- SUBCELLULAR LOCATION: NUCLEAR.

CC -1- TISSUE SPECIFICITY: THYROID, LUNG AND CNS. EXPRESSED IN RESTRICTED

CC REGIONS OF THE DEVELOPING BRAIN WITHIN THE DIENCEPHALON, IN PARTS

CC OF THE HYPOTHALAMUS AND NEUROHYPOPHYSIS, AND IN THE TELECEPHALON.

CC -1- SIMILARITY: BELONGS TO THE NK-2 FAMILY OF HOMEBOX PROTEINS.

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: X53858; CA37851.1; -; ALT-SEQ.

DR EMBL: D38035; BAA07231.1; ALT-SEQ.

DR PIR: S12002; S12002.

DR PDB: 1FTT; 282AN-96.

DR TRANSLAC: T00835; -;

DR PROSITE: PS00027; HOMEBOX_1; 1.

DR PROSITE: PS50071; HOMEBOX_2; 1.

DR PIRAM: PF00046; homeobox; 1.

RV Transcription regulation; Activator; Homeobox; DNA-binding;

KV Nuclear protein; 3D-structure.

FT DNA_BIND 161 220 HOMEBOX.

FT DOMAIN 234 244 POLY-GLY.

FT DOMAIN 247 254 POLY-GLN.

FT DOMAIN 295 304 POLY-ALA.

SO SEQUENCE 372 AA; 38554 MW; AE0F1572 CRC32;

TF1_RAT Length: 372 February 14, 2000 08:02 Type: P Check: 6608 ..

1 MSMSPKHTTP FSVSDILSPL EESYKVGME GGLGAPLAA YKOGAAPP

51 AAMQOHAVGH HGAVTAAHYH TAAGVPOLSH SAVGCGYNGN LGMSLPPY

101 QDTRNSASG PGWYGANPD RPAPSRFNG PASGMNSGM GGLSGDVS

151 KMAPLPSP RRRRVLFSG AOVYELERF KQKYLISAP REHLASMTL

201 TPQVKITFO NHRKMKRQA KDKAAQOOLQ ODSGGGGGG GAGACFQOQ

251 AQOQSPRVA VPVYKDGK COAGAPAPGA ASLOGHAQO AQOQAQAOA

301 AAAAISVSG GAGIGAHPGH OPGSAGSPD LAHHAASPA LOGGVSSLSH

351 LNSGSDYGA MSCSTLYGR TW

!!AA_SEQUENCE 1.0 STANDARD: PRT; 81 AA.

ID TXF8_DENAN

AC P01404;

DT 21-JUL-1986 (rel. 01. Created)

DT 01-FEB-1994 (rel. 28. Last sequence update)

DT 01-FEB-1994 (rel. 28. Last annotation update)

DE TOXIN F-VIII PRECURSOR (TOXIN TAZ) (TOXIN DAF8).

OS Dendroaspis angusticeps (Eastern green mamba).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Lepidosauria;

OC Squamata; Scleroglossa; Serpentes; Colubroidae; Elapidae; Elapinae;

OC Dendroaspis.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-VEINOM GLAND;

RA DUCANCEL F.;

RN Submitted (JUN-1990) to the EMBL/Genbank/DBJ databases.

RP SEQUENCE OF 22-81.

RC TISSUE-VEINOM;

RX MEDLINE: 74070661.

RA VILJOEN C.C., BOTES D.P.;

RT "Snake venom toxins. The purification and amino acid sequence of

RT toxin Taz from Dendroaspis angusticeps venom."

RL J. Biol. Chem. 249:3863-3872(1974).

CC -1- FUNCTION: ITS PHARMACOLOGICAL ACTION IS UNKNOWN. IT PROBABLY ACTS

CC SYNERGISTICALLY WITH OTHER VENOM COMPONENTS AS WHOLE VENOM IS VERY

CC TOXIC.

CC -1- MISCELLANEOUS: LD(50) IS >250 MG/KG BY SUBCUTANEOUS INJECTION.

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: X53409; CA37485.1; -;

DR PIR: A01675; TSEPZA.

DR PIR: S21299; S21299.

DR HSSP: P01416; INTX.

DR PROSITE: PS00272; SNAKE_TOXIN; 1.

DR PIRAM: PF00087; toxin; 1.

KV Venom; Toxin; Signal

FT SIGNAL 1 21

FT CHAIN 22 81

FT DISULFID 24 43 TOXIN F-VIII.

FT DISULFID 38 60 BY SIMILARITY.

FT DISULFID 62 73 BY SIMILARITY.

FT DISULFID 74 79 BY SIMILARITY.

FT CONFLICT 54 54 I -> V (IN REF. 2).

SO SEQUENCE 81 AA; 8849 MW; E0BFC94E CRC32;

TXF8_DENAN Length: 81 February 14, 2000 08:02 Type: P Check: 8703 ..

1 MKTLLTLV VTIYCLDAS TMIYSHKTP OPSATITCEE KTCYKRSYK

51 LPAIVAGRG GCPKREMLVA IHCERSDKCN E

!!AA_SEQUENCE 1.0 STANDARD: PRT; 153 AA.

ID UBC7_DROME

AC P52487;

DT 01-OCT-1996 (rel. 34. Created)

DT 01-OCT-1996 (rel. 34. Last sequence update)

DT 01-NOV-1997 (rel. 35. Last annotation update)

DE UBQUITIN-CONJUGATING ENZYME P2-18 KD (EC 6.3.2.19) (UBIQUITIN-PROTEIN

DE LIASE) (UBIQUITIN CARRIER PROTEIN).

GN UBC84D.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Insecta; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

RN [1]

RP SEQUENCE FROM N.A.

RC ROBIN C., MEDVECKY K.M., RUSSELL R.J., OAKESHOTT J.G.;

RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO OTHER
 CC PROTEINS (BY SIMILARITY).
 CC
 CC -1- CATALYTIC ACTIVITY: ATP + UBIQUITIN + PROTEIN LYSINE -
 CC AMP + PYROPHOSPHATE + PROTEIN N-UBIQUITYLATION.
 CC
 CC -1- PATHWAY: SECOND STEP IN UBIQUITIN CONJUGATION.
 CC
 CC -1- MISCELLANEOUS: A Cysteine residue is required for
 CC UBIQUITIN-THIOLESTER FORMATION (BY SIMILARITY).
 CC
 CC -1- SIMILARITY: BELONGS TO THE UBIQUITIN-CONJUGATING ENZYME FAMILY.
 CC STRONGEST: TO HUMAN UBC7.
 CC
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL: U51051; AAA01150.1; -
 CC DR HSSP: P15731; 20CE.
 CC DR ELYBASE: FBq0017456; UBC84D.
 CC DR PROSITE: PS00183; UBIQUITIN_CONJUGAT; 1.
 CC DR PFAM: PF00179; UQ.con; 1.
 CC DR Ubiquitin conjugation: Llgase.
 CC KW BINDING: 86 86 UBIQUITIN (BY SIMILARITY).
 CC FT SEQUENCE: 153 AA; 17709 MW; 41698C20 CRC32;
 CC
 CC UBC7_DROME Length: 153 February 14, 2000 08:02 Type: P Check: 1891 ..
 CC
 CC 1 MATRILITE LSDLYEAMS TLNIESSE SLMTGLIY PERAPIYKA
 CC 51 XRIINEPPO YEMPEKILF KTIYHPND EKGEVCLPIL STDNWKETTR
 CC 101 TQVQLALVA IVHNPPEHP LRSDLAEEV REHKFKMKA DEFTKNAEK
 CC 151 RPE
 CC
 CC !!AA_SEQUENCE 1.0
 CC ID UMUD_ECOLI STANDARD: PRT; 139 AA.
 CC AC P04153;
 CC DT 01-NOV-1986 (Rel. 03, Created)
 CC DT 15-JUL-1999 (Rel. 03, Last sequence update)
 CC DT 15-JUL-1999 (Rel. 38, Last annotation update)
 CC DE UMUD PROTEIN (EC 3.4.21.-) [CONTAINS: UMUD' PROTEIN].
 CC GN UMUD.
 CC OS Escherichia coli.
 CC OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC OC Escherichia.
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE: 85242678.
 CC RA KITAGAWA Y., AKASHI E., SHINAGAWA H., HORII T., OGAWA H., KATO T.;
 CC RT Structural analysis of the umu operon required for inducible
 CC RT mutagenesis in Escherichia coli.;
 CC RT Proc. Natl. Acad. Sci. U.S.A. 82:4336-4340(1985).
 CC RL [2]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE: 85242678.
 CC RA PERRY K.L., ELLIOTT S.J., MITCHELL B.B., MARSH L., WALKER G.C.;
 CC RT "umuC and umuB operons whose products are required for UV light-
 CC RT and chemical-induced mutagenesis: Umud, MucA, and LexA proteins share
 CC RT homology.";
 CC RT Proc. Natl. Acad. Sci. U.S.A. 82:4331-4335(1985).
 CC RL [3]
 CC RP SEQUENCE FROM N.A.
 CC RX STRAIN-K12 / MG1655;
 CC RX MEDLINE: 97426617.
 CC RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
 CC RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
 CC RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GODDEN M.A., ROSE D.J.,
 CC RA MAU B., SHAO Y.;
 CC RT "The complete genome sequence of Escherichia coli K-12.";

RL Science 277:1453-1474(1997).
 RN [4]
 RN RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE: 97061202.
 RA OSHIMA T., ALBA H., BABA T., FUJITA K., HAYASHI K., HONJO A.,
 RA IKEMOTO K., INABA T., ITOH T., KAJIHARA M., KAWAI K., KASHIMOTO K.,
 RA KIMURA S., KITAGAWA M., MAKINO K., MASUDA S., MIKI T., MIZOBUCHI K.,
 RA MORI H., MOTOMURA K., NAKAMURA Y., NASHIMOTO H., NISHIO Y., SATO N.,
 RA SANEI G., SEKI T., TAGAMI H., TAKEMOTO K., WADA C., YAMAMOTO Y.,
 RA YANO M., HORIUCHI T.;
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 12,728,0 min region on the linkage map.";
 RT DNA Res. 3:137-155(1996).
 RN [5]
 RN RP MUTAGENESIS.
 RA MEDLINE: 92318898.
 RA KOCH W.H., ENNIS D.G., LEVINE A.S., WOODGATE R.;
 RT "Escherichia coli umud mutants: DNA sequence alterations and umud
 RT cleavage.";
 RT Mol. Gen. Genet. 233:443-448(1992).
 RN [6]
 RN RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF UMUD'.
 RX MEDLINE: 96202734.
 RA PEAT T.S., FRANK E., McDONALD J.P., LEVINE A.S., WOODGATE R.,
 RA HENDRICKSON W.A.;
 RT "Structure of the umud' protein and its regulation in response to DNA
 RT damage.";
 RT Nature 380:727-730(1996).
 CC -1- FUNCTION: INVOLVED IN UV PROTECTION AND MUTATION. ESSENTIAL FOR
 CC INDUCED (OR SOS) MUTAGENESIS. MAY MODIFY THE DNA REPLICATION
 CC MACHINERY TO ALLOW BYPASS SYNTHESIS ACROSS A DAMAGED TEMPLATE.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S24 ALSO KNOWN AS THE
 CC UMUD/LEXA FAMILY.
 CC
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL: M10107; AAA24728.1; -
 CC DR EMBL: M13387; AAA98073.1; -
 CC DR EMBL: AE000216; AAC74267.1; -
 CC DR EMBL: D90752; BAA36030.1; -
 CC DR EMBL: D90753; BAA36038.1; -
 CC DR PIR: A03551; ZWECB.
 CC DR PIR: A23157; A23157.
 CC DR PDB: 1UMU; 01-ANG-96.
 CC DR PDB: 1AY9; 28-JAN-98.
 CC DR ECO2DBASE: A015.1; 6TH EDITION.
 CC DR ECOGENE: EG11057; UMUD.
 CC DR FRAY, PFO717; Peptidase_S24.1
 CC SOS mutagenesis; DNA repair; Autocatalytic cleavage; Hydrolase;
 CC KM Serine protease; 3D-structure.
 CC FT PROPEP 1 24
 CC FT CHAIN 25 139 UMUD' PROTEIN.
 CC FT SITE 24 25 CLEAVAGE (AUTO-).
 CC FT ACT_SITE 60 60 INVOLVED IN AUTO-CLEAVAGE.
 CC FT ACT_SITE 97 97 INVOLVED IN AUTO-CLEAVAGE.
 CC FT MUTAGEN 27 27 P->D: IN UMUD1; NON-CLEAVABLE.
 CC FT MUTAGEN 65 65 G->R: IN UMUD44; NON-CLEAVABLE.
 CC FT MUTAGEN 92 92 G->D: IN UMUD77; NON-CLEAVABLE.
 CC SQ SEQUENCE: 139 AA; 15063 MW; D0175DFA CRC32;
 CC
 CC UMUD_ECOLI Length: 139 February 14, 2000 08:02 Type: P Check: 8403 ..
 CC
 CC 1 MFIIPALR EITPPLFSD LVQCFSPSA ADVEGRID NQLLIQHSFA
 CC 51 TTFVASGDS MIDGISDGD LLIYSALTA SHDIIVANV DEFTVKKLO

101 LRPVOLLIM NSAVSPITIS SEDTLDVGV VHVYKAMR

11AA_SEQUENCE 1.0
ID US02_HSVK STANDARD; PRT: 303 AA.
AC 006240;
DT 01-OCT-1994 (Rel. 27, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE US1 PROTEIN.
GN US1.
OS Equine herpesvirus type 1 (strain Kentucky A) (EHV-1).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicelloviruses.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93079867.
RA BREEDEN C.A., YALAMANCHILI R.R., COLLE C.F. III, O'CALLAGHAN D.J.;
RT "Identification and transcriptional mapping of genes encoded at the
RT IR/US junction of equine herpesvirus type 1."
RL Virology 191:649-660(1992).
RN [2]
RP SEQUENCE OF 1-209 FROM N.A.
RX MEDLINE: 92263758.
RA COLLE C.F. III, FLOWERS C.C., O'CALLAGHAN D.J.;
RT "Open reading frames encoding a protein kinase, homolog of
RT glycoprotein gX or pseudorabies virus, and a novel glycoprotein map
RT within the unique short segment of equine herpesvirus type 1."
RL Virology 188:345-357(1992).
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES US2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M87497; AAA46069.1; -
DR PIR: A44215; A44215. E -> G (IN REF. 2).
FT CONFLICT 31 31
FT MOD_RES 31 31
SQ SEQUENCE 303 AA: 33502 MW: 839D6CID CRC32;
US02_HSVK Length: 303 February 14, 2000 08:02 Type: P Check: 5996 ..

1 MGVLIVVT VVDRKALPN SSIDVDGHLW EFLRQCFVL ASEPLGPIY
51 VRSDLYRFS SSLLTPKAC RPIYTRGDT AALDRNGVY YHEDRMGSI
101 EMLSVLSGYN HLNSSLITNQ PYHLWVGLAA DLCKPFDLI PGKRMVYAE
151 IADFEKSMQ PPFCVCKLFE TTPWTVEHN HPLKLAAGG EDTVVGECGF
201 SKHSSNSLVH PPTVNRVIYA VVDPARLKEI PAFGRPLPRR RPSESGMRAP
251 RRRSRAPAPA RSTAAATPP RPPGDRAPAA RRAGDVYME RLWCVGERT
301 STR

11AA_SEQUENCE 1.0
ID VANS_ENTFC STANDARD; PRT: 384 AA.
AC 006240;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE SENSOR PROTEIN VANS (EC 2.7.3.-) (VANCOMYCIN RESISTANCE PROTEIN VANS)
DE (VANCOMYCIN HISTIDINE PROTEIN KINASE).
GN VANS.
OS Enterococcus faecium (Streptococcus faecium).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC Enterococcus.
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN-BM4147; TRANSPOSON-TN1546;
RX MEDLINE: 93106944.
RA ARTHUR M., MOLINAS C., DEPARDIE F., COUVALIN P.;
RT "Characterization of Tn1546, a Tn3-related transposon conferring
RT glycopeptide resistance by synthesis of depsipeptide peptidoglycan
RT precursors in Enterococcus faecium BM4147."
RL J. Bacteriol. 175:117-127(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BM4147;
RX MEDLINE: 92210502.
RA ARTHUR M., MOLINAS C., COUVALIN P.;
RT "The Vans-Vanr two-component regulatory system controls synthesis of
RT depsipeptide peptidoglycan precursors in Enterococcus faecium
RT BM4147."
RL J. Bacteriol. 174:2582-2591(1992).
CC -1- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM VANS/VANR.
CC ACTIVATES THE TRANSCRIPTION OF VANH, VANA AND VANX IN RESPONSE TO
CC VANCOMYCIN WHICH RESULTS IN VANCOMYCIN RESISTANCE. VANS ACTIVATES
CC VANR BY PHOSPHORYLATION.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: TO OTHER PROKARYOTIC SENSOR TRANSDUCTION HISTIDINE
CC KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M87297; AAA65954.1; -
DR EMBL: M68510; AAA24788.1; -
DR PIR: PF00512; Signal; 1.
KW Sensory transduction; Transferase; Kinase; Transmembrane; Cell wall;
KW Antibiotic resistance; Phosphorylation; Plasmid.
FT TRANSMEM 21 41
FT TRANSMEM 76 96
FT MOD_RES 164 164
FT MOD_RES 164 164
SQ SEQUENCE 384 AA: 43915 MW: 89EA01F7 CRC32;
VANS_ENTFC Length: 384 February 14, 2000 08:02 Type: P Check: 3601 ..

1 MYIKLNKNK DYSLERKLY MYVAIVVA IVFVLYIRSM IKGKLDWIL
51 SIENKDYDN HLDAMKLYOY SIRNNDIFI YVAIVSILI LCRVMSKPA
101 KYFDELNGI DVLQNEKQ IELSAEMDVN EQKNTLKRI LERQEDATL
151 AQRKNDVVM YLADIKITPL TSIGYLSL DEARDMPVDQ KAKYVHITID
201 KAYRLEQLID EFFEITRYNL QTITLTKTHI DLYTMLYOMT DEYFPLSAH
251 GKQAVIHAPE DLYSGDPDK LARVENNIUK NAAVSEDSN IIDITGSLG
301 DVVSIEFKNT GSIPKDLAA IEFKTYRLDN ARSSDTGGAG LGALAKELI
351 VOHGOIYAE SNDNYTTFV ELPAMDLDV KRRS

11AA_SEQUENCE 1.0
ID VES2_HPV03 STANDARD; PRT: 383 AA.
AC P36778;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE REGULATORY PROTEIN E2.
GN E2.
OS Human papillomavirus type 3.
OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomaviruses.
RN [1]
RP SEQUENCE FROM N.A.

```

RA MEDLINE 94265501.
RA DELIUS H., HOFMANN B. ;
RT "Primer-directed sequencing of human papillomavirus types." ;
RL Curr. Top. Microbiol. Immunol. 186:13-31(1994).
CC -1- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
CC IT BINDS TO THE E2R RESPONSE ELEMENT (5'-ACCNANNNGG-3') PRESENT
CC IN MULTIPLE COPIES IN THE REGULATOR REGION. IT CAN EITHER
CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2R'S POSITION
CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
CC REPLICATION.
CC -1- SUPERNIT. BINDS DNA AS A DIMER.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC ..
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
CC EMBL: X74462; CAA52472.1; -.
CC DR PIR: S36552; S36552.
CC DR HSSP: P17383; IDHM.
CC DR PFAM: PF00508; E2_N; 1.
CC DR PFAM: PF00511; E2_C; 1.
CC KW Early protein; Transcription regulation; Activator; DNA-binding;
CC Trans-acting factor; DNA replication; Repressor; Nuclear protein.
CC SO SEQUENCE 383 AA; 44246 MW; CEBDIDEA CRC32.
VE2_HP003 Length: 383 February 14, 2000 08:02 Type: P Check: 4819 ..

1 METLANLDV CODKILELYE KDSQKLEDDI MHWOLMRLEQ ALTYKAREGC
51 LTHGHDQVP PLSVTKAKAR SALEVHVSQ QLOSHAHMD PWTLRDTSRE
101 MMDTVPRKMC KRGGLVEVR YDQDENKAMC YQWMEILVQ NYTDMDWYAV
151 AGLVSHGEL YMHGQKTFY VKRKDARYV GDTGWDVHV GKVYIHDSF
201 DPAVSSTREIP AGRPLYACT QAPQAQVCA SECPQKQKQ LFTVYGEQQ
251 QQQQQQQQQQ HQTQPAQPT ERAKQPLDID TRDRDITCP HPIGHRSDD
301 CVPYHLRCD PNCLEKCFYR LNRGNKLVS RTSSTWRMSC ESENOCAYVT
351 IWTYSYGQRE AFLSTVKVP GIOVYLGHMS MFT

11AA_SEQUENCE 1.0
ID VE2_HP010 STANDARD: PRG. 376 AA.
AC P36781;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE REGULATORY PROTEIN E2.
GN E2.
OS Human papillomavirus type 10.
OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE 94265501.
RX DELIUS H., HOFMANN B. ;
RT "Primer-directed sequencing of human papillomavirus types." ;
RL Curr. Top. Microbiol. Immunol. 186:13-31(1994).
CC -1- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
CC IT BINDS TO THE E2R RESPONSE ELEMENT (5'-ACCNANNNGG-3') PRESENT
CC IN MULTIPLE COPIES IN THE REGULATOR REGION. IT CAN EITHER
CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2R'S POSITION
CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
CC REPLICATION.
CC -1- SUPERNIT. BINDS DNA AS A DIMER.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC ..
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
CC EMBL: X74462; CAA52472.1; -.
CC DR PIR: S36552; S36552.
CC DR HSSP: P17383; IDHM.
CC DR PFAM: PF00508; E2_N; 1.
CC DR PFAM: PF00511; E2_C; 1.
CC KW Early protein; Transcription regulation; Activator; DNA-binding;
CC Trans-acting factor; DNA replication; Repressor; Nuclear protein.
CC SO SEQUENCE 383 AA; 44246 MW; CEBDIDEA CRC32.

```

```

CC      REPLICATION.
CC      CC -1- SUBUNIT: BINDS DNA AS A DIMER.
CC      CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC      CC -----
CC      CC This swiss-prot entry is copyright. It is produced through a collaboration
CC      CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC      CC the European Bioinformatics Institute. There are no restrictions on
CC      CC use by non-profit institutions as long as its content is in no
CC      CC modified and this statement is not removed. Usage by and for commer
CC      CC entities requires a license agreement (See http://www.isb-sib.ch/annou
CC      CC or send an email to license@isb-sib.ch).
CC      CC -----
CC      DR EMBL: X74465; CAAS2492.1; -
CC      DR PIR: S36535; S36535.
CC      DR HSSP: P17883; IDHM.
CC      DR PRAM: PF00508; E2L_N; 1.
CC      DR PRAM: PF00511; E2L_C; 1.
CC      KW Entry protein; transcription regulation; activator; DNA-binding;
CC      KW trans-acting factor; DNA replication; Repressor; Nuclear protein.
CC      SQ SEQUENCE 376 AA; 43003 MW; 799683D3 CRC32;
CC      VE2_HPV10 Length: 376 February 14, 2000 08:02 Type: P Check: 273 ..
CC
CC      1 METLANRDLA QODKMLEYE KDSKLEDOI THMILLRYEN ALLYKARECG
CC      51 LTHIGHQVPP PLSTVTRAKAR NAIDVHALQ OLOESAVANE PWTLRDTSRE
CC      101 MMDIAPKGCW KKRGITVEYR YDGDSEKAKC YVOMELLYQ NYSDBRWKV
CC      151 PKGVSYEGLY YTHENNMIYV VNFDDDACVY GETGWEYHV GKVVIHDAF
CC      201 DPKVSYREIS TPGVCTSTNT TPASTQAVQV ASEGEQKQK RLEAVDGOHQ
CC      251 QQRQSGKDSI OKAAERAGQV VDSQRTSLCD TRSAHPVHRP SDPCDCAPIH
CC      301 LRQDPKSLKC FRYRLHGRK KLYRSRSSTW RMCSESEQA AFVTLWYSD
CC      351 TQTEFLANV KVPKGIOVL GYMSIF
CC
CC      11AA-SEQUENCE 1.0
CC      ID VE2_HPV18 STANDARD; PRT; 365 AA.
CC      AC P06760;
CC      DT 01-JUN-1998 (Rel. 06, Created)
CC      DT 01-APR-1999 (Rel. 25, Last sequence update)
CC      DT 15-JUL-1998 (Rel. 36, Last annotation update)
CC      DE REGULATORI PROTEIN E2.
CC      GN E2.
CC      OS Human papillomavirus type 18.
CC      OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.
CC      RN [1]
CC      RP SEQUENCE FROM N.A.
CC      RX MEDLINE: 87283882.
CC      RA COLE S.T.; DAVOS O.;
CC      RA "Nucleotide sequence and comparative analysis of the human
CC      RA papillomavirus type 18 genome. Phylogeny of Papillomaviruses and
CC      RA repeated structure of the E6 and E7 gene products."
CC      J. Mol. Biol. 193:599-608(1987).
CC      [2]
CC      RP REVISION TO 90.
CC      RP MEDLINE: 93197132.
CC      RA MEISSNER J.;
CC      RA "Pacl is a single cut enzyme for HPV-18."
CC      Nucleic Acids Res. 21:1041-1041(1993).
CC      CC -1- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
CC      IT BINDS TO THE EAR RESPONSE ELEMENT (5'-ACCCNNNNNGG-3') PRESENT
CC      CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
CC      CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2AR'S POSITION
CC      CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
CC      CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
CC      CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
CC      CC REPLICATION.
CC      CC -1- SUBUNIT: BINDS DNA AS A DIMER.
CC      CC -1- SUBCELLULAR LOCATION: NUCLEAR.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X05015; GAA28667.1; ALT_SEQ.
DR PIR: D26251; W2RL18.
DR HSSP: P17383; IDHM.
DR PFAM: PF00508; E2_N; 1.
DR PFAM: PF00511; E2_C; 1.
KW Early protein; transcription regulation; Activator; DNA-binding;
KW Trans-acting factor; DNA replication; Repressor; Nuclear protein.
SQ SEQUENCE 365 AA; 41294 MW; 187C1FE CRC32;
VE2_HPV18 Length: 365 February 14, 2000 08:02 Type: P Check: 5006 ..
1 MOTPKETISE RLSCVODKII DHYENDSKDI DSOIQWOLI RMENALFFAA
51 REHGIOTLNH QVPAVYNISK SKAKHAIELQ MALQGLAOSA YKTEDWTLOD
101 TCEELMTEP THCFKGGQI VQYFPGNND NCMYVANDS VYMTDAGTW
151 DKATCVSHR GLYVKEGYN TFEIERKSEC EKYGTGTWE VHFNNVIDC
201 NDSMSTSD TVSATQLVKQ LQHTSPYSS TVSGTATY GQSAATRPQ
251 HGLAEKHC GVPNPLGAA TPFGNKKRRK LCGNTTPII HLGDRNSLK
301 CLRYRKHS DHYRDISHTW HMGAGNERT GLTYTYSSE TQTKFLNTV
351 AIPDSQIIV GYTM
!!A_SEQUENCE 1.0 STANDARD: PRT: 376 AA.
ID VE2_HPV28
AC P50771;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE REGULATORY PROTEIN E2.
GN E2.
OS Human Papillomavirus type 28.
OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.
RN [1]
RP SEQUENCE FROM N.A.
RA DELIUS H.;
RL Submitted (OCT-1995) to the EMBL/Genbank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U31783; AAA79425.1; ..
DR HSSP: P17383; IDHM.
DR PFAM: PF00508; E2_N; 1.
DR PFAM: PF00511; E2_C; 1.
SQ SEQUENCE 376 AA; 43234 MW; 466EDC47 CRC32;
VE2_HPV28 Length: 376 February 14, 2000 08:02 Type: P Check: 1816 ..
1 METLANRLDV CODKMLELYE KDSNKLEDQI MHMOLMRVEN ALLYKARECG
51 LTHIGHQVVP PLSTVKAKAR SAIEVHALL OLOESAYAD SWLTDTSRE
101 MMDIVPKCW KRGVTVFVR YDGDETRSMC YVHMDIFTO NYSDDKWKV

151 AGHVSIEGLY YIHGEQTFY VKFKDAVYV GETGKEVHV GKVIHHHAF
201 DPVSTREIP AAGPLTSDT TKASTETSVG ATEGPOOKRO RLETLNMEQO
251 QROYPQTPST QTERASQPL DVTRTSDCT TCPYVGHPS DPDCAPVHL
301 KGDNCNCKCF RYRLHKGKRR LYCKTSTSTW WSCSENOAA FVLIWTSYS
351 QRNELSTVK VPPGIQVILG HNSMFV
!!A_SEQUENCE 1.0 STANDARD: PRT: 391 AA.
ID VE2_HPV2A
AC P25482;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE REGULATORY PROTEIN E2.
GN E2.
OS Human Papillomavirus type 2a.
OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.
RN [1]
RP SEQUENCE FROM N.A.
RA HIRSCH-BEHNAM A.; DELIUS H.; DE VILLIERS E.M.;
RT "A comparative sequence analysis of two human papillomavirus (HPV)
RT types 2a and 57.";
RL Virus Res. 18:81-98(1990).
CC -1- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
CC IT BINDS TO THE E2E RESPONSE ELEMENT (5'-ACCCNNNNNGGT-3') PRESENT
CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDENT OF E2E'S POSITION
CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
CC REPLICATION.
CC -1- SUBUNIT: BINDS DNA AS A DIMER.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X55964; -; NOT_ANNOTATED_CDS.
DR PIR: S15617; S15617.
DR HSSP: P17383; IDHM.
DR PFAM: PF00508; E2_N; 1.
DR PFAM: PF00511; E2_C; 1.
KW Early protein; transcription regulation; Activator; DNA-binding;
KW Trans-acting factor; DNA replication; Repressor; Nuclear protein.
SQ SEQUENCE 391 AA; 43233 MW; B3DE7AF6 CRC32;
VE2_HPV2A Length: 391 February 14, 2000 08:02 Type: P Check: 734 ..
1 METLANRDA COETILELYE KDSNKLEDQI KHMVQVLEN VMLFKARECG
51 MTRVGTAVP ALYSKAKAC QALIEVQLAQ TIMQSAVSTE AMTLRDTCLE
101 MMDAPPKCW KKGQSVLYK FDGSSDRDRI YTSMGFIYVQ DITLDSMHV
151 PGQVDELGLY YVHGVAVNY VDFGDESLEY GYGTGWEVHV AGVYIHHISA
201 SVSTQASAS DDEPLSPIRT AVSPVAPVYA ASASTAGR AAPYQALCS
251 AQAFTSPPAK KQRYVGOQH PPDDSTRYVG EGEVECTYNR SISDSKRDP
301 RMGHDDTDSV PVHLRGDAN CLKCFRYVQ KHKQVLYARV SSTHWAQGN
351 GDTAFVTLW YTSVEQRTTF LTRVSIKGL IALPGWSAF V

```

11AA_SEQUENCE 1.0      STANDARD:      PRT:      372 AA.
ID_V02_HPV31
AC_P1783;
DT_01-AUG-1990 (Rel. 15, Created)
DT_01-AUG-1990 (Rel. 15, last sequence update)
DT_15-JUL-1998 (Rel. 36, last annotation update)
DE_REGULATORY PROTEIN E2.
GN_E2.
OS_Human Papillomavirus type 31.
RN_Viruses; dsDNA viruses, no RNA stage: Papovaviridae; Papillomavirus.
RP_SEQUENCE FROM N.A.
RX_MEDLINE: 89299478.
RT_GOLDSPRING M.D.: DISTINGUISH D. TEMPLE G.F., LORINCZ A.T.:
RT_neoplasm-associated virus."
RT_virology 171:306-311(1989).
RN_STRUCTURE BY NMR OF 291-372.
RX_MEDLINE: 96194130.
RX_LING H., PETROS A.M., MEADOWS R.P., YOON H.S., EGAN D.A., WALTER K.,
RA_HOLZMAN T.F., ROBINS T., FESIK S.W.:
RT_Solution structure of the DNA-binding domain of a human
RT_papillomavirus E2 protein: evidence for flexible DNA-binding
RT_regions."
RL_Biochemistry 35:2095-2103(1996).
CC_FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
CC_IT BINDS TO THE E2E RESPONSE ELEMENT (5'-ACNNNNNNNGT-3'), PRESENT
CC_IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
CC_ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2E'S POSITION
CC_WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
CC_BY SERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
CC_INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
CC_REPLICATION.
CC_-1- SUBUNIT: BINDS DNA AS A DIMER.
CC_-1- SUBCELLULAR LOCATION: NUCLEAR.
CC_-1- SUBCELLULAR LOCATION: NUCLEAR.
CC_THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
CC_BETWEEN THE SWISS INSTITUTE OF BIOFORMATICS AND THE EMBL OUTSTATION -
CC_THE EUROPEAN BIOFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
CC_USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
CC_MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
CC_ENTITIES REQUIRES A LICENSE AGREEMENT (See http://www.isb-sib.ch/announce/
CC_OR SEND AN EMAIL TO license@isb-sib.ch).
CC-----
DR_EMBL: J04343; AAA6953.1.
DR_PIR: D32444; W2M31.
DR_PDB: 1DHW; 07-DEC-96.
DR_PFM: PFO0508; E2_N; 1.
DR_PFM: PFO0511; E2_C; 1.
KW_Early protein; transcription regulation; Activator; DNA-binding;
KW_Tumor-acting factor; DNA replication; Repressor; Nuclear protein;
KW_3D-structure.
SQ_SEQUENCE 372 AA: 42104 MW; B34B2C1 CRC32;
VE2_HPV31 Length: 372 February 14, 2000 08:02 Type: P Check: 9656
1 METLSRLNVAQCDKLTLEHYE NDSKRLCDHI DYMKHIRLEC VLMYKARENG
51 IHSINQVAP ALVSARKARL QAIELQMLE TLNNTEYKNE DMTMOOTSDE
101 LYLAFTGCL KKHGYTVEQV FDEADVNTNH YTNMFYILC IDQOCTVVG
151 QVNCNGIYV HEGHITYEVN FTEAKKYGK GKMEVHAGQ QVIVPESEV
201 SSDSEIFAGI VTKLPTANNT TTSNKTICAL GKSEGVRAIT TSKRPRTBP
251 EHRTHHPNK ILRGDSVDVY NGCVISAAC TNGRAVSCP ATPIIHLNG
301 DANIKGLRY RLSKYKQLYE QVESTHMYNC TDGKHKNALV TLVIISTSR
351 DDFLNTYKIP NTVSVSTGYM TI

```

```

ID      ID     PRT:    368 AA.
IIAA_SEQUENCE 1.0
VE2_HPV45 STANDARD;
AC       AC        P36794;
DT       DT         01-JUN-1994 (Rel. 29, Created)
PT       PT          01-JUL-1994 (Rel. 29, last sequence update)
DE       DE           15-JUL-1998 (Rel. 36, last annotation update)
CN       CN            REGULATORY PROTEIN E2.
OS       OS             Human papillomavirus type 45.
CC       CC              Vtuseset; dsDNA vtuseses, no RNA stage; Papovaviridae; Papillomavirus.
RN       RN               [1]
RP       RP                SEQUENCE FROM N.A.
RX       RX                 MEDLINE; 9426501.
RA       RA                  DELIUS H., HOFMANN B.;
RF       RF                   Primer-directed sequencing of human papillomavirus types.";
RL       RL                    Curr. Top. Microbiol. Immunol. 186:13-31(1994).
RC       RC                     -1 FUNCTION: E2 REGULATE VIRAL TRANSCRIPTION AND DNA REPLICATION. IT BINDS TO THE E2R RESPONSE ELEMENT ('5'-ACCNNNNGG(-3') PRESENT IN MULTIPLE COPIES IN THE REGULATOR REGION. IT CAN EITHER ACTIVATE OR SUPPRESS TRANSCRIPTION DEPENDING OF E2R'S POSITION WITH RESPECT TO PROXIMAL PROMOTER ELEMENTS. REPRESSON OCCURS BY STERICALLY HIDING THE ASSEMBLY OF THE TRANSCRIPTION INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA REPLICATION.
CC       CC                      -1 SUBUNIT: BINDS DNA AS A DIMER.
CC       CC                       -1 SUBCELLULAR LOCATION: NUCLEAR.
-----
CC       CC                        This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch).
-----
DR       DR      EMBL; X74479; CAA52576.1; .
DR       DR      PIR; S36564; S36564
DR       DR      HSHP; P17383; 1DHW.
DR       DR      PRAM; PF00508; E2.N: 1.
DR       DR      PRAM; PF00511; E2.C: 1.
KW       KW      Early protein; Transcription regulation; Activator; DNA-binding; Trans-acting factor; DNA replication; Repressor; Nuclear protein.
SQ       SQ        Sequence 368 aa; 41946 MW; 7D4610A6 CRC32;.
VS_2_HP45 Length: 368 February 14, 2000 08:02 Type: p Check: 5828 ..
1 MKQTPRESI SERLSADLK IDDHYENDSK DINSQISYQ LIRENALIF
51 TANEHGITYK NHQVPPINI SKSAKHAIIE LOMALKGLAQ SKYNNEWTL
101 QDTCEELIMNT EPSOCFFKKGG KTVHVFEDGN KDNCNMYYVM DSIIYTETGG
151 IMDKTAAGYS YWGVIYYIKDG DTFIVYDFKS ECERYGSNSMT WEVOYGANVI
201 DCNDMSMGTS DDIVSATDIY ROLQHASTIT PKIASVGTPK PHIQTPATKR
251 PROCGLTEGH HGAVTHVHN PLICSTSNM KRKYCSGNT FTIHLLGDCK
301 NSUKCLRFL RKYADHYSEI SSTHWGTGCN KNILITIVI NSEYOBNFTL
351 DVVTIPNSVQ ISGYMI
IIAA_SEQUENCE 1.0 STANDARD; PRT: 358 AA.
ID VE2_HPV51
AC P26547;
DT 01-AUG-1992 (Rel. 23, Created)
PT 01-AUG-1992 (Rel. 23, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DE REGULATORY PROTEIN E2.
CN GN E2.
OS OS Human papillomavirus type 51.
CC CC Vtuseset; dsDNA vtuseses, no RNA stage; Papovaviridae; Papillomavirus.

```


RP SEQUENCE FROM N.A.
 RX MEDLINE: 91303675.
 RA LINGU O., CRUM C.P., SILVERSTEIN S.J.:
 RT "Biologic properties and nucleotide sequence analysis of human
 RT papillomavirus type 51".
 RL J. Virol. 65:4216-4225(1991).
 CC -1- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
 CC IT BINDS TO THE E2E RESPONSE ELEMENT (5'-ACGNNNNNGGT-3') PRESENT
 CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
 CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING ON E2E'S POSITION
 CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
 CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
 CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
 CC REPLICATION.
 CC -1- SUBUNIT BINDS DNA AS A DIMER.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M62877; -; NOT_ANNOTATED.CDS.
 DR PIR: B40415; M2ML51.
 DR HSSP: P17383; 1DHM.
 DR PFAM: PF00508; E2.N; 1.
 DR PFAM: PF00511; E2.C; 1.
 KW Early protein; Transcription regulation; Activator; DNA-binding;
 KW Trans-acting factor; DNA replication; Repressor; Nuclear protein.
 SQ SEQUENCE 358 AA: 40908 MW: E0C694E CRC32:
 VE2_HP51 Length: 358 February 14, 2000 08:02 Type: P Check: 1816 ..
 1 METICHLNLY COEKILDCYE LDSKLDYQI NYWTLRLRYEA AMFYAREBN
 51 LRTINHOVVP ATTYSKQKAC QALEMHALQ SLKSDIYNE PPTMRETCYE
 101 LMCVAPROCF KKGITIVYI FDGNKNAMD YTSMKFIYI DNDKRWKTING
 151 NVDYTGIIYT VNSKREYVO FKDEAKIYA QQNEVYMYGT VITCEYVSS
 201 TCSDALSTT TVEOLSNPT TNPITCVGA KEAOTOORR OHLPEPDSST
 251 ISPLSVNTN NOIHCSSGST NTGHSATQ TAFIVHLKGD TNLCKCFYR
 301 FTKKGLYKN VSTWHTSN TKTGIVIVE DSAHQRETFI KTIKPPSVT
 351 LSLGIMTL
 11AA_SEQUENCE 1.0 STANDARD; PRT; 368 AA.
 ID VE2_HP52
 AC P36796;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 13-JUL-1998 (Rel. 36, Last annotation update)
 DE REGULATORY PROTEIN E2.
 GN E2.
 OS Human papillomavirus type 52.
 CC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94265501.
 RA DEIDUS H., HOFMANN B.:
 RT "Primer-directed sequencing of human papillomavirus types.";
 RL Curr. Top. Microbiol. Immunol. 186:13-31(1994).
 CC -1- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
 CC IT BINDS TO THE E2E RESPONSE ELEMENT (5'-ACGNNNNNGGT-3') PRESENT
 CC IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
 CC ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING ON E2E'S POSITION
 CC WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS

CC BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
 CC INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
 CC REPLICATION.
 CC -1- SUBUNIT BINDS DNA AS A DIMER.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X74481; CA525588.1; -
 DR PIR: S36576; S36576.
 DR HSSP: P17383; 1DHM.
 DR PFAM: PF00508; E2.N; 1.
 DR PFAM: PF00511; E2.C; 1.
 KW Early protein; Transcription regulation; Activator; DNA-binding;
 KW Trans-acting factor; DNA replication; Repressor; Nuclear protein.
 SQ SEQUENCE 368 AA: 41739 MW: 918AE9D9 CRC32:
 VE2_HP52 Length: 368 February 14, 2000 08:02 Type: P Check: 4491 ..
 1 MESIPARLNA VOEKILDCYE ADSNDINAQI EHMKLTREMC VLFYKAKELG
 51 ITHIGHQVVP PPAVSKAKAC QALELOLAL ALNKQYSTD GWTLOQTSLE
 101 MMAEPQKVF KKHGYTIVQ YDNDKNMTD YTNMREIYLL GECECTIVEG
 151 QVDYGYIYV CDGEKIYVK FSDAKQYCV TGWSEVHGG QVIVCPASVS
 201 SNEVSTTEA VHLCTETSST SAVSVGAKD HLPQPKRRR PDVDSNTRK
 251 YPNNLKGGQ SYDSTRGLV TATECTNKR VAHTCTAPI IHLKGDNSL
 301 KCLRVRKTH KSLYQIIST WHWTSNCTN NNLGIVITLY SDETRQROFL
 351 KTVKIRPVO VLOQVMSL
 11AA_SEQUENCE 1.0 STANDARD; PRT; 317 AA.
 ID VEAA_BP22
 AC O03544;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last annotation update)
 DE EAA PROTEIN.
 GN EAA.
 OS Bacteriophage P22.
 CC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Podoviridae.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94018622.
 RA WILF D.L., HO Y.S., POWERS S., ROSENBERG M.:
 RT "The int genes of bacteriophages P22 and lambda are regulated by
 RT different mechanisms".
 RL Mol. Microbiol. 9:261-271(1993).
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: L06296; AAC18882.1; -
 DR PIR: S35282; S35282.
 SQ SEQUENCE 317 AA: 35719 MW: 05933D59 CRC32:
 VEAA_BP22 Length: 317 February 14, 2000 08:02 Type: P Check: 7758 ..

1 MTIKERIE LFVKSPIENG LTRGEOMELA RIALASDAE TVRINKFSG
51 TCVTLEQOPN AADDVAVIIP LYAAPVPER ERIRREHAEM SOKTFEDVGP
101 VGPLKHLKSE ALFAADPSD PLEKADMOFL LMDAQRKNGI SDEFITIRAI
151 EKLEINKRKQ WPEPKDPEPR LIKEQPESEV VPECPALP YAOVAVADL
201 YALCMOSEEV VTTPDPEKA TWINNYSCT CQETVKLER LQALAGNSP
251 VIPGWMSCS ERMNDDESK PLAIFTKCL GQGMFVATYD DQGFEDYWG
301 MEIGVSHWM QLPDPPL
!!AA_SEQUENCE 1.0 STANDARD; PRT; 177 AA.
ID VGG_BP64
AC P03644;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE MAJOR SPIKE PROTEIN (G PROTEIN) (GPG).
GN Bacteriophage G4.
OS Bacteriophage G4.
OC Viruses; ssDNA viruses; Microviridae; Microvirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 79053264.
RA GODSON G.N., BARRELL B.G., STADEN R., FIDDES J.C.;
RL "Nucleotide sequence of bacteriophage G4 DNA.";
RL Nature 276:236-247(1978).
RM [2]
RN SEQUENCE OF 1-81 FROM N.A.
RX MEDLINE: 78248733.
RA SIMS J., DRESSLER D.;
RL "Site-specific initiation of a DNA fragment: nucleotide sequence of
RL the bacteriophage G4 negative-strand initiation site.";
RL Proc. Natl. Acad. Sci. U.S.A. 75:3094-3098(1978).
RM [3]
RN X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
RX MEDLINE: 96217890.
RA MCKENNA R., BOWMAN B.R., IIAG L.L., ROSSMANN M.G., FANE B.A.;
RL "Atomic structure of the degraded procapsid particle of the
RL bacteriophage G4: induced structural changes in the presence of
RL calcium ions and functional implications.";
RL J. Mol. Biol. 256:736-750(1996).
CC -1- FUNCTION: MAJOR SPIKE COMPONENT. INVOLVED IN THE ATTACHMENT TO
CC THE BACTERIAL HOST.
CC -1- SUBUNIT: THE VIRION IS COMPOSED OF 60 COPIES EACH OF THE F, G, AND
CC J PROTEINS, AND 12 COPIES OF THE H PROTEIN. THERE ARE 12 SPIKES
CC WHICH ARE EACH COMPOSED OF 5 G AND ONE H PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: J02454; AAA32324.1;
DR EMBL: V00657; CA24020.1;
DR EMBL: M25080; AAA32328.1;
DR PIR: A04252; ZGBPG4.
DR PDB: 1GFF; 03-APR-96.
KW Coat protein; 3D-structure.
FT CONFLICT 27 A -> T (IN REF. 2).
FT SEQUENCE 177 AA; 18820 MW; 700A8E43 CRC32;
VGG_BP64 Length: 177 February 14, 2000 08:02 Type: P Check: 4484

1 MFQKFSKIN APINSTOLAA TKTPAVAPV LSVPLUSST ILLNATTIV
51 THSGICHAV RIDETNPIN HALSTAGSIS NVPADMIFA IREVDAGV

101 PFAVALYDV YPIETFNK AISEKDAVTI DSHPTVGNV VYAGIMLSN
151 AMTASTISGV LSVNOVNEA TVLOPLK
!!AA_SEQUENCE 1.0 STANDARD; PRT; 372 AA.
ID VGLI_HSV2H
AC P13261;
DT 01-JAN-1980 (Rel. 13, Created)
DT 01-JAN-1980 (Rel. 13, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE GLYCOPROTEIN 1.
GN G1 OR US7.
OS Herpes simplex virus (type 2 / strain HG52).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 87111457.
RA MCGEOCH D.J., MOSS H.W.M., MCNAB D., FRAME M.C.;
RL "DNA sequence and genetic content of the HindIII 1 region in the
RL short unique component of the herpes simplex virus type 2 genome:
RL identification of the gene encoding glycoprotein G, and evolutionary
RL comparisons.";
RL J. Gen. Virol. 68:19-38(1987).
RM [2]
RN SEQUENCE FROM N.A.
RA DOLAN A.;
RL Submitted (FEB-1997) to the EMBL/Genbank/DBJ databases.
CC -1- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1 AND
CC 2 GH, GB, GC, GD, GI AND GE.
CC -1- SIMILARITY: TO OTHER HERPESVIRUSES GLYCOPROTEINS 1, TO VZV GP1V,
CC AND TO PRV GP63.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X04798; CA28485.1;
DR EMBL: Z86099; CAB06714.1;
DR PIR: F43674; F43674.
KW GLYCOPROTEIN.
FT CARBOHYD 156 156 POTENTIAL.
FT CARBOHYD 169 169 POTENTIAL.
FT CARBOHYD 175 175 POTENTIAL.
FT CARBOHYD 243 243 POTENTIAL.
FT SEQUENCE 372 AA; 39558 MW; E7FA7C46 CRC32;
VGLI_HSV2H Length: 372 February 14, 2000 08:02 Type: P Check: 501
1 MPGSLDGLA ILGLWCAG LVAGPVS LVSLSVDAGA VGPGFVEED
51 LVYGEHLHFV GACVPHINTY DGIELFHYR LGNHCRVYH VYTLACPRR
101 PAVAFICRS THRAHPAYP TLEIGLAROP LNRVATATD YAGLYLRFV
151 VGSATNARSF VLVALSANG TLYVNGSDYG SCDDPAOLPS APRLPSSVY
201 TPGASRPTTP RTTPPSSPR DPTAPADTG TPAASGEIA PPNSTRASE
251 SRRLTVAOV IQAIPASII AFVLGSCIC FHRRCORRYR RPRGOIYNG
301 GVSADVNEAA MARLGAELRS HPNTPPKRR RSSSTTMS LTSIAESEP
351 GPVVLVSVP RPRSGTAPQ EV
!!AA_SEQUENCE 1.0 STANDARD; PRT; 366 AA.
ID VGLM_HSVGA
AC Q01017;

DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE GLYCOPROTEIN M.
 GN 39.
 OS Herpesvirus salmieri (strain 11).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 92333688.
 RA ALBRECHT J.-C., NICHOLAS J., BILLER D., CAMERON K.R., BIESINGER B.,
 RA NEWMAN C., WITTMANN S., CRAIXON M.A., COLEMAN H., FLECKENSTEIN B.,
 RA HONESS R.W.;
 RT "Primary structure of the herpesvirus salmieri genome.";
 RT J. Virol. 66:5047-5058(1992).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: TO OTHER HERPESVIRUSES GLYCOPROTEIN M.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC -----
 CC EMBL: X61346; CAA5662.1; -
 DR PIR: A36810; ODBENS.
 DR PIR: P01528; Herpes_glycop. 1.
 KW Transmembrane; Glycoprotein.
 FT TRANSMEM 18 34 POTENTIAL.
 FT TRANSMEM 87 103 POTENTIAL.
 FT TRANSMEM 152 168 POTENTIAL.
 FT TRANSMEM 214 230 POTENTIAL.
 FT TRANSMEM 236 252 POTENTIAL.
 FT TRANSMEM 274 290 POTENTIAL.
 FT TRANSMEM 305 321 POTENTIAL.
 FT DOMAIN 343 351 ARG/LYS-RICH (BASIC).
 FT CARBOHYD 57 57 POTENTIAL.
 SQ SEQUENCE 366 AA; 42182 MW; 14D99C2B CRC32;
 VGLM_HSVSA Length: 366 February 14, 2000 08:02 Type: P Check: 7071 ..
 1 MKMASRSDTF MLRTWIOQLV LFVIMFMSA ILPIASVVG LGFCYPPNL
 51 VDSLSNLTL RNAKHILPT LFLAPBLFV YITWSVYDL ASAIYYVGA
 101 LAIQARKR LRSMTLQTM INLVGSHNL FIGIAMMTL QLFHYLSYK
 151 HVMIAFIY LFHCLSYMT LSLVSRNSPK WSVLMECHI PKOSLSITIL
 201 DYGRPLCVNM YLSLALLENL VESLGFMAI GNSFTLVSD TVLASINLIY
 251 VLTFEWMYT EMFLDYELK QFGFYLGVS GSILILPVL RYEAFFVSAN
 301 LKRTVAANIA MIPACVIM MRLFRYSQ VPKPNSYTP LPRKRRRO
 351 KODQOLIME TSDEEL
 11AA_SEQUENCE 1.0
 ID VIRB_AGR5 STANDARD; PRT; 237 AA.
 AC P17798;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 01-NOV-1990 (Rel. 16, Last annotation update)
 DE VIRB8 PROTEIN.
 GN Agrobacterium tumefaciens.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Agrobacterium.
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE: 90318324.
 RA KUDDAU G.A., DE VOS G., OMEN J., MCCAFFREY G., ZAMBRYSKI P.;
 RT "The vir operon of Agrobacterium tumefaciens pTiC58 encodes 11 open
 RT reading frames.";
 RT Mol. Gen. Genet. 221:256-266(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 90301800.
 RA ROGONSKIY P.M., POWELL B.S., SHTRASU K., LIN T.-S., MOREL P.,
 RA ZYRINK E.M., STECK T.R., KADO C.I.;
 RT Molecular characterization of the vir region of Agrobacterium
 RT tumefaciens: complete nucleotide sequence and gene organization of
 RT the 28.63-kbp region cloned as a single unit.
 RT Plasmid 23:85-106(1990).
 CC -1- FUNCTION: VIRB PROTEINS ARE SUGGESTED TO ACT AT THE BACTERIAL
 CC SURFACE AND THERE PLAY AN IMPORTANT ROLE IN DIRECTING T-DNA
 CC TRANSFER TO PLANT CELLS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC -----
 CC EMBL: X53264; CAA37361.1; -
 DR EMBL: J03320; AAA91596.1; -
 DR PIR: S12348; B8AG58.
 KW Crown gall tumor; Plasmid.
 FT CONFLICT 21 22 MISSING (IN REF. 1).
 FT CONFLICT 128 129 SA -> TR (IN REF. 2).
 SQ SEQUENCE 237 AA; 26294 MW; CD671958 CRC32;
 VIRB_AGR5 Length: 237 February 14, 2000 08:02 Type: P Check: 5994 ..
 1 MKGSYVALV ARETLAEHYK EYEAFOYRA KSARRLSKVI AAVAILAVLG
 51 NVQAQFTAT MVLIRLIPV YIMIRPDGTV DSEYVSRLP AJOEAVVNA
 101 SLMEYVRURE SYDADTAOYA YDLVSNFSAF MVRQNYOOFF NYPNPTSPV
 151 ILGKHGRLEV EHTASNDVTP GVOQIRYKRT LIYDGKAPMA STWATVRYE
 201 KVTSLPGRLR LFNPGGLVVT SYOTSEDTVS NAGHSEP
 11AA_SEQUENCE 1.0
 ID VIRB_AGR6 STANDARD; PRT; 230 AA.
 AC P09781;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE VIRB8 PROTEIN.
 GN Agrobacterium tumefaciens.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Agrobacterium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 88186901.
 RA WARD J.E., AKIYOSHI D.E., REGIER D., DATTA A., GORDON M.P.,
 RA NESTER E.W.;
 RT "Characterization of the virB operon from an Agrobacterium
 RT tumefaciens Ti plasmid.";
 RT J. Biol. Chem. 263:5804-5814(1988).
 RN [2]
 RP REVISIONS.
 RX MEDLINE: 90170994.
 RA WARD J.E., AKIYOSHI D.E., REGIER D., DATTA A., GORDON M.P.,
 RA NESTER E.W.;
 RL J. Biol. Chem. 265:4768-4768(1990).

CC -1- FUNCTION: VIRB PROTEINS ARE SUGGESTED TO ACT AT THE BACTERIAL
CC SURFACE AND THERE PLAY AN IMPORTANT ROLE IN DIRECTING T-DNA
CC TRANSFER TO PLANT CELLS.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.sib.ch/announce/>
CC or send an email to license@isb.sib.ch).

CC EMBL: J03216; -; NOT ANNOTATED_CDS.

DR PIR: I28621; B8AG65.

KW Crown gall tumor; Plasmid.

SO SEQUENCE 230 AA; 25362 MW; 44A8A66C CRC32;

VIB8_AGRt6 Length: 230 February 14, 2000 08:02 Type: P Check: 2646 ..

1 MLVARESLAE HYKEVEAFOT ARAKSARRUS KIINAVALIA ILGNVAQAFa
51 IATWPLSLRL VPYLMIRAD GTVDSEVSIS RLPAQOEAV VNASLMEYVR
101 LRESYDADTA QYAYDLVSNE SAPTVRODYO QEFNYPNSS POYLGRGR
151 VEVEHIASND VTPSTOQIRY KRLVYDGKM PVSTWTATV RYKVTSLPG
201 RLRLNPAQL VTSYOTSED TVSNVGGAP

11AA_SEQUENCE 1.0 STANDARD; PRT; 230 AA.

ID VIB8_AGRt9

AC P05357;

DT 01-NOV-1988 (Rel. 09, Created)

DT 01-AUG-1991 (Rel. 19, Last sequence update)

DT 01-MAY-1992 (Rel. 22, Last annotation update)

DE VIRB8 PROTEIN.

GN Agrobacterium tumefaciens.

OS Plasmid pRi15953.

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

CC Rhizobiaceae; Agrobacterium.

CC [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 88247765.

RA THOMPSON D.V., MELCHERS L.S., IDLER K.B., SHILPEROORT R.A.,
RA HOOKKAS P.J.J.;

RT "Analysis of the complete nucleotide sequence of the Agrobacterium
RT tumefaciens virB operon."

RT Nucleic Acids Res. 16:4621-4636(1988).

CC -1- FUNCTION: VIRB PROTEINS ARE SUGGESTED TO ACT AT THE BACTERIAL
CC SURFACE AND THERE PLAY AN IMPORTANT ROLE IN DIRECTING T-DNA
CC TRANSFER TO PLANT CELLS.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.sib.ch/announce/>
CC or send an email to license@isb.sib.ch).

CC EMBL: X06826; CAA29978.1; ALT_INT.

DR PIR: S00784; B8AG55.

KW Crown gall tumor; Plasmid.

SO SEQUENCE 230 AA; 25408 MW; 870AF438 CRC32;

VIB8_AGRt9 Length: 230 February 14, 2000 08:02 Type: P Check: 2826 ..

1 MLVARESLAE HYKEVEAFOT ARAKSARRUS KIINAVALIA ILGNVAQAFa
51 IATWPLSLRL VPYLMIRAD GTVDSEVSIS RLPAQOEAV VNASLMEYVR
101 LRESYDADTA QYAYDLVSNE SAPTVRODYO QEFNYPNSS POYLGRGR

151 VEVEHIASND VTPSTOQIRY KRLVYDGKM PVSTWTATV RYKVTSLPG
201 RLRLNPAQL VTSYOTSED TVSNVGGAP

11AA_SEQUENCE 1.0 STANDARD; PRT; 147 AA.

ID VIB1_AGRt9

AC P13461;

DT 01-JAN-1990 (Rel. 13, Created)

DT 01-JAN-1990 (Rel. 13, Last sequence update)

DT 01-JAN-1990 (Rel. 13, Last annotation update)

DE T-DNA BORDER ENDONUCLEASE VIRID1 (EC 3.1.1.-).

GN VIRID1.

OS Agrobacterium rhizogenes.

OC Plasmid pRi14B.

CC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

CC Rhizobiaceae; Agrobacterium.

CC [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 89039712.

RA HIRAYAMA T., MURANAKA T., OKAWA H., OKA A.;

RT "Organization and characterization of the virC gene from
RT Agrobacterium rhizogenes."

RT Mol. Gen. Genet. 213:229-237(1988).

CC -1- FUNCTION: TUMOR FORMATION BY A. TUMEFACIENS INVOLVES THE TRANSFER
CC AND INTEGRATION OF A DEFINED SEGMENT (T-DNA) OF TI PLASMID DNA
CC INTO THE PLANT NUCLEAR GENOME. THE VIRD OPERON ENCODES A SITE-
CC SPECIFIC ENDONUCLEASE THAT CLEAVES AT A UNIQUE SITE WITHIN BOTH
CC 24 BP DIRECT REPEATS FLANKING THE T-DNA.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.sib.ch/announce/>
CC or send an email to license@isb.sib.ch).

CC EMBL: X12867; CAA31350.1; -.

DR PIR: S06883; S06883.

KW Hydrolyase; Nuclease; Endonuclease; Crown gall tumor; Plasmid;

CC T-DNA.

SO SEQUENCE 147 AA; 16005 MW; 414FDDA0 CRC32;

VID1_AGRt9 Length: 147 February 14, 2000 08:02 Type: P Check: 2761 ..

1 MSOGSKPTSS DIAINORVGA TVEGFRVST RLRSAYESF SHQARLLGUS

51 DSAIRAVARA RIGGFLEIDA ETRHKEMALI LSIGTLSSNI ALLSAVAEN

101 PTMDELELRA ERINGESFA DDDGLRLSLI SVSRRRIDGC SMKDSLS

11AA_SEQUENCE 1.0 STANDARD; PRT; 147 AA.

ID VIB1_AGRt9

AC P18551; P06520;

DT 01-JAN-1988 (Rel. 06, Created)

DT 01-NOV-1990 (Rel. 16, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE T-DNA BORDER ENDONUCLEASE VIRID1 (EC 3.1.1.-).

GN VIRID1.

OS Agrobacterium tumefaciens.

OC Plasmid pRiC58.

CC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

CC Rhizobiaceae; Agrobacterium.

CC [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 90301800.

RA ROGOWSKY P.M., POWELL B.S., SHIRASU K., LIN T.-S., MOREL P.,
RA ZYRARIAN E.M., STECK T.R., KADO C.I.;

RT "Molecular characterization of the vir regulon of Agrobacterium
RT tumefaciens: complete nucleotide sequence and gene organization of
RT the 28.63-kbp regulon cloned as a single unit."

RT Plasmid 23:85-106(1990).

```

RN PRELIMINARY SEQUENCE FROM N.A.
RX HADIVA M., CLOST T.J., TAIT R.C., KADO C.I.;
RT "Identification of pTIC58 plasmid-encoded proteins for virulence in
RT Agrobacterium tumefaciens."
RT Proc. Natl. Acad. Sci. U.S.A. 82:2669-2673(1985).
RN [3]
RN SEQUENCE FROM N.A.
RN WANG K., HERREIRA-ESTRELLA A.H., VAN MONTAGU M.M.;
RN Submitted (JUL-1990) to the EMBL/GenBank/DBD databases.
RA -1- FUNCTION: TUMOR FORMATION BY A TUMEFACIENS INVOLVES THE TRANSFER
CC AND INTEGRATION OF A DEFINED SEGMENT (T-DNA) OF TI PLASMID DNA
CC INTO THE PLANT NUCLEAR GENOME. THE VIRID OPERON ENCODES A SITE-
CC SPECIFIC ENDONUCLEASE THAT CLEAVES AT A UNIQUE SITE WITHIN BOTH
CC 24 BP DIRECT REPEATS FLANKING THE T-DNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J03320; AAA91603.1;
DR EMBL: M11311; AAA98365.1; ALT_SEQ.
DR EMBL: M33673; AAA22110.1;
DR PIR: A22666; A22666.
DR PIR: S11838; S11838.
KM Hydrolyase; Nuclease; Endonuclease; Crown gall tumor; Plasmid;
SO SEQUENCE 147 AA; 16130 MW; 7187146F CRC32;

VID1_AGRT5 Length: 147 February 14, 2000 08:02 Type: P Check: 2312 ..

1 MSQSRPTSS DIAYNORECV KYEGKRVST RLSAEYEF SHQARLLGLS
51 DSMAIRAVAR RIGGFLEIDA ETRHREMAIL QSIGTLSSNT AALLSAVAEN
101 PTMDLEALRA ERIAFKSF A DDLGLRSIL SVSRRRIDGC SLKDAL

11AA_SEQUENCE 1.0
ID VID1_AGRT6 STANDARD; PRT; 147 AA.
AC P06667;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-OCT-1989 (Rel. 12, Last annotation update)
DE T-DNA BORDER ENDONUCLEASE VID1 (EC 3.1.-.-).
GN VIR1.
OS Agrobacterium tumefaciens.
OG Plasmid pTiA6NC.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
CC Rhizobiaceae; Agrobacterium.
RN [1]
RN SEQUENCE FROM N.A.
RN MEDLINE: 87028239.
RA NESTER E.W., FORSTER S.G., YOUNG C., ALBRIGHT L.M., GORDON M.P.,
RT "The virD operon of Agrobacterium tumefaciens encodes a site-specific
RT endonuclease."
RT Cell 47:471-477(1986).
RN [2]
RN SEQUENCE FROM N.A.
RN MEDLINE: 88038822.
RA JAVASMAL R.K., VELUTHAMBI K., GELVIN S.B., SLIGHTOM J.L.;
RT "Double-stranded cleavage of T-DNA and generation of single-stranded
RT T-DNA molecules in Escherichia coli by a virD-encoded
RT border-specific endonuclease from Agrobacterium tumefaciens."
RT J. Bacteriol. 169:5035-5045(1987).
CC -1- FUNCTION: TUMOR FORMATION BY A TUMEFACIENS INVOLVES THE TRANSFER
CC AND INTEGRATION OF A DEFINED SEGMENT (T-DNA) OF TI PLASMID DNA
CC INTO THE PLANT NUCLEAR GENOME. THE VIRID OPERON ENCODES A SITE-

```

```

CC SPECIFIC ENDONUCLEASE THAT CLEAVES AT A UNIQUE SITE WITHIN BOTH
CC 24 BP DIRECT REPEATS FLANKING THE T-DNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M41762; AAA98369.1;
DR EMBL: M41989; AAA22113.1;
DR PIR: A29826; A29826.
DR PIR: A25063; A25063.
KM Hydrolyase; Nuclease; Endonuclease; Crown gall tumor; Plasmid;
SO SEQUENCE 147 AA; 16215 MW; 305D4C9C CRC32;

VID1_AGRT6 Length: 147 February 14, 2000 08:02 Type: P Check: 4145 ..

1 MSKHRTYSS ETRAINQHSRL NVGEFKVSA RLSAEYEF SYQARLLGLS
51 DSMAIRAVAR RIGGFLEIDA HTRKREMAIL QSIGTLSSNV SMLSAVAED
101 PRSDLEAVRD ERIAFGEAFA ALDGLRSIL SVSRRRIDGC SLKDAL

11AA_SEQUENCE 1.0
ID VF1_NPVOP STANDARD; PRT; 374 AA.
AC O10330;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE VERY LATE EXPRESSION FACTOR 1.
GN VF-1.
OS Oryza pseudoturgata multicausid polyhedrosis virus (OpNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
CC Nucleopolyhedrovirus.
RN [1]
RN SEQUENCE FROM N.A.
RN MEDLINE: 97271300.
RA ARENS C.H., RUSSELL R.R., FUNK C.J., EVANS J., HARWOOD S.,
RT "The sequence of the Oryza pseudoturgata multicausid polyhedrosis virus genome."
RT Virology 229:381-399(1997).
CC -1- FUNCTION: INVOLVED IN VERY LATE GENE ACTIVATION (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U75930; AAC59079.1;
DR PIR: P00589; P00589.
KM Transcription regulation.
SO SEQUENCE 374 AA; 43165 MW; C560A0FD CRC32;

VF1_NPVOP Length: 374 February 14, 2000 08:03 Type: P Check: 889 ..

1 MDGLGVNET TENDMKARIQ SASREHVED LAIDROCTP DEVANDSILMS
51 KYMFPPAP TILSKYSRL IKILFSLVEE ADLQNPAYL DREDDSVFQ
101 HLVSPELC KRLLEIRSVT KETLQILTFE YTNANMLPEF KIPRMVMPR
151 DREKTIRES ENFMKKNAI DTILNFDISK IKLMSDYVH DRGLIRGAIV
201 FCIMIGTMR INEAROLVD DLNVLIRKGR LRSDTIGLR KRSRKNLNN

```

251 IKTRPLELAR EYARRPVL QISKSTSTPF KDFRRLDEA GYEMERPRSN
301 MIRHLYSSNL YNSGVPLQKV ARLMNHESPA STKPYLNKNY FDESSDEES
351 GGNRRDSSTG SSANSSSLTY QTGD

11AA-SEQUENCE 1.0 STANDARD: PRT: 186 AA.
ID VMA2_HRSVA

AC P29792;
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DE MATRIX GLYCOPROTEIN M2.
GN M2.
OS Bovine respiratory syncytial virus (strain A51908) (BRS).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Pneumovirinae; Pneumovirus.
RN [1]
RP SEQUENCE FROM N.A.

RA ZAMORA M., SAMAL S.K.;
RX MEDLINE: 92185490.
RT "Sequence analysis of M2 mRNA of bovine respiratory syncytial virus
RT obtained from an F-M2 distronic mRNA suggests structural homology
RT with that of human respiratory syncytial virus.";
RL J. Gen. Virol. 73:737-741(1992).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
CC or send an email to license@sdb-sdb.ch).

DR EMBL: M82816; AAA42805.1; -
DR PIR: J01482; WMNZBA.
KW Matrix protein: Envelope protein: Glycoprotein.
FT CARBOHYD 54 POTENTIAL.
FT CARBOHYD 89 POTENTIAL.
SQ SEQUENCE 186 AA: 21351 MW: CCA02395 CRC32;

VMA2_HRSVA Length: 186 February 14, 2000 08:03 Type: P Check: 6337 ..

1 MSRRNPCKYE IRGHCLNGKK CHFSHNYFEM PPHALLVRON FMLNKLKSM
51 DRNDITLSEI SGAALDRTE EYALGVIVL ESYLSISNNI TKOSACVAMS
101 KLAEINND IKRLRNKEVP TSPKIRIYNT VISIYDSNKR NTKOTIHLK
151 RLPADVLKKT IKNTIDIHNE INGNNOGDIN VDEONE

11AA-SEQUENCE 1.0 STANDARD: PRT: 194 AA.
ID VMA2_HRSVA

AC P04545;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE MATRIX GLYCOPROTEIN M2 (ENVELOPE-ASSOCIATED 22 KD PROTEIN).
GN M2.
OS Human respiratory syncytial virus (strain A2).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Pneumovirinae; Pneumovirus.
RN [1]
RP SEQUENCE FROM N.A.

RA ELANGO N., SARAKH M., VENKATESAN S.;
RX MEDLINE: 85237684.
RT "RNA sequence of three respiratory syncytial virus genes encoding
RT two nonstructural proteins and a 22k structural protein.";
RL J. Virol. 55:101-110(1985).

[2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 85135082.
RA COLLINS P.L., WERTZ G.W.;

RT "The envelope-associated 22k protein of human respiratory syncytial
RT virus: nucleotide sequence of the mRNA and a related
RT polytranscript.";
RL J. Virol. 54:65-71(1985).

[3]
RP SEQUENCE FROM N.A.
RX MEDLINE: 95266253.
RA CONNORS M., CROWE J.E. JR., FIRESTONE C.Y., MURPHY B.R., COLLINS P.L.;
RT "A cold-passaged, attenuated strain of human respiratory syncytial
RT virus contains mutations in the F and L genes.";
RL J. Virol. 208:478-484(1995).

[4]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97187925.
RA CROWE J.E. JR., FIRESTONE C.Y., WHITEHEAD S.S., COLLINS P.L.;
RA MURPHY B.R.;
RT "Acquisition of the ts phenotype by a chemically mutagenized cold-
RT passaged human respiratory syncytial virus vaccine candidate results
RT from the acquisition of a single mutation in the polymerase (L)
RT gene.";
RL Virus Genes 13:269-273(1996).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
CC or send an email to license@sdb-sdb.ch).

DR EMBL: M11486; AAB59860.1; -
DR EMBL: U50362; AAB86655.1; -
DR EMBL: U50363; AAB86677.1; -
DR EMBL: U63644; AAC5971.1; -
DR PIR: A04034; WMNZ22.
KW Matrix protein: Envelope protein: Glycoprotein.
FT CARBOHYD 89 POTENTIAL.
FT CARBOHYD 191 POTENTIAL.
SQ SEQUENCE 194 AA: 22154 MW: 5997358C CRC32;

VMA2_HRSVA Length: 194 February 14, 2000 08:03 Type: P Check: 9828 ..

1 MSRRNPCKYE IRGHCLNGKR CHFSHNYFEM PPHALLVRON FMLNKLKSM
51 DKSIDITLSEI SGAALDRTE EYALGVIVL ESYIGSISNNI TKOSACVAMS
101 KLTELNSDD IKRLRDNEL NSPKIRIYNT VISIYESNRK NTKOTIHLK
151 RLPADVLKKT IKNTLDIHS ITINPKEST VSDINDHAKN NDTT

11AA-SEQUENCE 1.0 STANDARD: PRT: 186 AA.
ID VMA2_ORSVW

AC O84132;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE MATRIX GLYCOPROTEIN M2.
GN M2.
OS Ovine respiratory syncytial virus (strain MSU 83-1578) (ORSV).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Pneumovirinae; Pneumovirus.
RN [1]
RP SEQUENCE FROM N.A.

RA ALANSARI H.M., FORGIETER L.N.D.;
RX MEDLINE: 95088607.
RT "Molecular cloning and sequence analysis of the phosphoprotein,
RT nucleocapsid protein, matrix protein and 22k (M2) protein of the
RT ovine respiratory syncytial virus.";
RL J. Gen. Virol. 75:3597-3601(1994).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U02510; AAA51645.1; -
KW Matrix protein; Envelope protein; Glycoprotein.
FT CARBOHYD 54
FT CARBOHYD 89
SQ SEQUENCE 186 AA; 21133 MW; CD7955FA CRC32;

VMA2_ORSW Length: 186 February 14, 2000 08:03 Type: P Check: 6559

1 MSRRRRCAYE IRGHCLNGRK CHFSHYFEM PPHALLVRON EMLNKLKEM
51 DSSNLTSEI SGAEEDRTE EYALGVIVL ESYLGAVNNI TKRSACVAMS
101 KLGEINSDO IKGLRNKLP TSPKRIYNT VISYDSNKR NPKOTIHLK
151 RLPADVLRKT IKNTIDTINE INVNNPSDIG VNEONE

11AA_SEQUENCE 1.0
ID VMEM_CVB STANDARD; PRT: 106 AA.
AC P37989;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DE 11.4 KD MEMBRANE PROTEIN (ORF 3).
OS Chrysanthemum virus B (CVB).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92013948.
RA LEVAY K., ZAVRIEV S.;
RT "Nucleotide sequence and gene organization of the 3'-terminal region
RT of chrysanthemum virus B genomic RNA."
RL J. Gen. Virol. 72:2333-2337(1991).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: TO ORF3 PROTEIN FROM POTEVIRUSES AND TO THE 14 KD
CC PROTEIN FROM BSMV RNA 2BETA.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; S60150; CAB31868.1; -
DR PIR; J01248; J01248.
DR PFAM; PF01307; Plant_vir_prot; 1.
KW Transmembrane.
FT TRANSMEM 10
FT TRANSMEM 72
SQ SEQUENCE 106 AA; 11435 MW; CPA2CE4A CRC32;

VMEM_CVB Length: 106 February 14, 2000 08:03 Type: P Check: 566

1 MRLTPPPHT KVLVAIAGL SIYASLTYS RNTLPQVGDH SHLLPBGVY
51 KQGTITVYG GPRKLSLEG GFLPYQPPWF LVLLISAIF LLSGRGHR

101 VCGQCH

11AA_SEQUENCE 1.0
ID VMEM_LVX STANDARD; PRT: 108 AA.
AC P27331;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE HYPOTHETICAL 11.8 KD PROTEIN (ORF 3).
OS Lily virus X.

CC Viruses; ssRNA positive-strand viruses, no DNA stage; Potexvirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90218039.
RA MEMELINK J., VAN DER VLOOT C.I.M., LINTHORST H.J.M.,
RA DERKS A.F.L.M., ASJES C.J., BOL J.F.;
RT "Homologies between the genomes of a carlavirus (lily symptomless
RT virus) and a potexvirus (lily virus X) from lily plants."
RL J. Gen. Virol. 71:917-924(1990).
CC -1- SIMILARITY: TO ORF3 PROTEIN FROM OTHER POTEVIRUSES AND TO 12 KD
CC PROTEIN FROM CARLAVIRUSES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X15342; CA33395.1; -
DR PFAM; PF01307; Plant_vir_prot; 1.
KW Transmembrane.
FT TRANSMEM 9
FT TRANSMEM 29
SQ SEQUENCE 108 AA; 11767 MW; 6EBDF03 CRC32;

VMEM_LVX Length: 108 February 14, 2000 08:03 Type: P Check: 8138

1 MRLTPPPHT KPLIAYVVG TLAFAVLILT RNTLPHTGDN LMSLPHGCTY
51 CDGTRIRYNG GPRSHVPEL PAKSMALITY VAILLALHFS CLTRHVRHC
101 VICHTTSG

11AA_SEQUENCE 1.0
ID VMEM_PVMR STANDARD; PRT: 109 AA.
AC P17527;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE 12 KD MEMBRANE PROTEIN (ORF 3).
OS Potato virus M (Strain Russian) (PVM).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Carlavirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89293091.
RA RUPASOV V.V., MOROZOV S.Y., KANYUKA K.V., ZAVRIEV S.K.;
RT "Partial nucleotide sequence of potato virus M RNA shows similarities
RT to potexviruses in gene arrangement and the encoded amino acid
RT sequences."
RL J. Gen. Virol. 70:1861-1869(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91116326.
RX ZAVRIEV S.K., KANYUKA K.V., LEVAY K.E.;
RT "The genome organization of potato virus M RNA."
RL J. Gen. Virol. 72:9-14(1991).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: TO ORF3 PROTEIN FROM POTEVIRUSES AND TO THE 14 KD
CC PROTEIN FROM BSMV RNA 2BETA.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; D14449; BA03341.1; -
DR EMBL; X53062; CA37234.1; -
DR PIR; PND003; MWYIP2.

DR PIR: S21603; S21603.
 DR PIR: C54333; C54333.
 DR PFAM: PF01307; Plant_vir_prot; 1.
 KW Transmembrane. 10 30 POTENTIAL.
 FT TRANSMEM 73 93 POTENTIAL.
 SQ SEQUENCE 109 AA; 11907 MW; E3BEC370 CRC32;
 VMEW_PVWR Length: 109 February 14, 2000 08:03 Type: P Check: 3415 ..
 1 MFLTPPDEFT KYLSAALGV SLALVLLLI RSTLPVGDH DNLPHGMY
 51 RDGKSVFYNN SPGRINSIEA RKAPLQGPW AIYVLLVLLI WASHKLGPN
 101 CRACAGSHT
 11AA_SEQUENCE 1.0 STANDARD; PRT; 149 AA.
 ID VP8_BPPH6
 AC P07579;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 01-JAN-1990 (Rel. 13, Last annotation update)
 DE MAJOR NUCLEOCAPSID PROTEIN (P8 PROTEIN).
 GN Bacteriophage phi-6.
 OS Bacteriophage phi-6.
 OC Viruses; dsRNA viruses; Cystoviridae; Cystovirus.
 RN [1]
 RP MEDLINE: 86144085.
 RA MCGRAN I., MINICH L., FRANGIONE B.;
 RT "Nucleotide sequence of the small double-stranded RNA segment of
 bacteriophage phi 6: novel mechanism of natural translational
 control.";
 RL J. Virol. 58:142-151(1986).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M12921; AAA32358.1; -
 DR PIR: A23368; VMBPF6.
 KM Nucleocapsid.
 SQ SEQUENCE 149 AA; 16004 MW; 23F29489 CRC32;
 VP8_BPPH6 Length: 149 February 14, 2000 08:03 Type: P Check: 6252 ..
 1 MLBPVRAAA VPAIESAIA TPGVSRRA AIGSKVSPSA IIAVKSMPV
 51 VAGITLTAQIG STGYDAVQOL LENBEVAEM LKLSKRADE IOPDTICLUG
 101 QYREBEELVE DAARVVGWS NLIRLQALE LDIKTYGLKM QLNMGYRS
 11AA_SEQUENCE 1.0 STANDARD; PRT; 297 AA.
 ID VPRP_SALDU
 AC P24417;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE VIRULENCE GENES TRANSCRIPTIONAL ACTIVATOR.
 GN VSDA OR SPVR.
 OS Salmonella dublin.
 OC Plasmid PSD12.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 91251759.
 RA KRAUSE M., ROUDIER C., FIERER J., HARWOOD J., GUINNY D.;
 RT "Molecular analysis of the virulence locus of the Salmonella dublin
 RT

RT Plasmid PSD12.";
 RT Mol. Microbiol. 5:307-316(1991).
 CC -1- FUNCTION: POSITIVE REGULATOR FOR THE PLASMID-ENCODED VIRULENCE
 CC FACTORS SPVA, SPVB AND SPVC
 CC -1- SUBCELLULAR LOCATION: CYTOSOL; CYTOSOL;
 CC -1- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
 CC REGULATORS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X56727; CAA40047.1; -
 DR PIR: S15213; S15213.
 DR PROSITE: P500044; HTH_LYSR_FAMILY; 1.
 DR PFAM: PF00126; HTH_L1; 1.
 KW Plasmid; Virulence; Transcription regulation; Activator; DNA-binding.
 FT DNA_BIND 21 40 H-T-H MOTIF (POTENTIAL).
 SQ SEQUENCE 297 AA; 33849 MW; DA84490B CRC32;
 VPRP_SALDU Length: 297 February 14, 2000 08:03 Type: P Check: 7061 ..
 1 MFLINKKKK IFTIMETGS FSIATSVLXI TRTPLSRVIS GLERELKQRL
 51 FIRKNGTLIP TEFAQIYRK VKSHYFELHA LEQELGPTGK TKOLEIIFDE
 101 IYPSIAKNI ISALTISGOK TITMGAAVNS QIIELCOTN NCIVISARYV
 151 FHRRESLVCT SVEGVAWLF PKFFFLGKP DINRLAGIPV LREKAKNN
 201 LDTIHFEEQ TLGITNPAPS FDNVLFSSL YRLOQGLAML LIPVRCRL
 251 GSTDHAIHI KVALCTSLY YPTKRETPD YRKAKILIQ ELKOSTE
 11AA_SEQUENCE 1.0 STANDARD; PRT; 297 AA.
 ID VPRP_SALDU
 AC P13041;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE VIRULENCE GENES TRANSCRIPTIONAL ACTIVATOR.
 GN MKAC OR SPVR OR VAGA.
 OS Salmonella typhimurium.
 OC Plasmid 96 kb virulence pex102.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-1275 WILD TYPE;
 RX MEDLINE: 90016881.
 RA FOLINGER G.D., BAIRD G.D., WILLIAMSON C.M., LAX A.J.;
 RT "Nucleotide sequence of a plasmid gene involved in the virulence of
 RT salmonellas.";
 RL Nucleic Acids Res. 17:7983-7983(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 90060335.
 RA TAIRA S., RHEIN M.;
 RT "Molecular organization of genes constituting the virulence
 RT determinant on the salmonella typhimurium 96 kilobase pair plasmid.";
 RL FEBS Lett. 257:274-278(1989).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 92041614.
 RA CALDWELL A.L., GULIG P.A.;
 RT "The Salmonella typhimurium virulence plasmid encodes a positive
 RT regulator of a plasmid-encoded virulence gene.";
 RL J. Bacteriol. 173:7176-7185(1991).
 RN [4]

RP SEQUENCE OF 1-10.
 RX MEDLINE: 91244158.
 RA TAIRA S., BAUMANN M., RIKKONEN P., SUKUPOLVI S., RHEIN M.;
 RT "Amino-terminal sequence analysis of four plasmid-encoded virulence-
 associated proteins of *Salmonella typhimurium*.";
 RT FEMS Microbiol. Lett. 61:319-323(1991).
 CC -1- FUNCTION: POSITIVE REGULATOR FOR THE PLASMID-ENCODED VIRULENCE
 CC FACTORS SPV, SPVB, AND SPVC.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
 CC REGULATORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: X16111; CAA34244.1; .
 CC PIR: S06089; S06089.
 CC PIR: S06670; S06670.
 CC PROSITE: PS00044; HTH_LYSR_FAMILY; 1.
 CC PRAM: PF00126; HTH_1; 1.
 CC Plasmid: Virulence: Transcription regulation: Activator; DNA-binding.
 CC DNA BIND 21 40 H-T-H MOTIF (POTENTIAL).
 CC SEQUENCE 297 AA: 33835 MW: 765F2867 CRC32;
 SO
 VRRP-SALTY Length: 297 February 14, 2000 08:03 Type: P Check: 7032 ..
 1 MDPLINKKK IFITMETOS FSATSVLYI TRTPLSRVIS DLRELRKRL
 51 FIRKNGTLIP TEFAQTIRK VKSHYIFLHA LEOEIQPTGR TKOLEIITDE
 101 IYPSLEKNLI ISALITSGOK TNMGRAVNS QIIIELCQTN NCYISARNY
 151 FHRESLYCRT SVEGGMLEFI PKKFLCGRP DINRLAGTVP LFHESAKNFN
 201 LDITHEFEQ TLTITNPAFS EDNDLFSSL YRLQGLAML LIPVAVCRAL
 251 GLSTDHALHI KGVALCTSLY YPRKRETPD YRKAKRLIOO ELKOSTF
 11AA-SEQUENCE 1.0
 ID WN11.CHICK STANDARD: PRT: 354 AA.
 AC P49339;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE WNT-11 PROTEIN PRECURSOR.
 GN WNT11 OR WNT-11.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
 OC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 95298011.
 RA TANDA N., OHUCHI H., YOSHIOKA H., NOJI S., NOHNO T.;
 RT "A chicken Wnt gene, Wnt-11, is involved in dermal development.";
 RT Biochem. Biophys. Res. Commun. 211:123-129(1995).
 CC [2]
 CC SEQUENCE FROM N.A.
 RX MEDLINE: 96063018.
 RA TANDA N., KAWAKAMI Y., SAITO T., NOJI S., NOHNO T.;
 RT "Cloning and characterization of Wnt-4 and Wnt-11 cDNAs from chick
 RT embryo.";
 RT Dev. Biol. 167:277-281(1995).
 CC -1- FUNCTION: MAY PLAY A ROLE IN THE FORMATION OF DERMAL STRUCTURE,
 CC BOTH LIMB AND FEATHER BUDS. IS LIKELY TO SIGNAL OVER ONLY FEW CELL
 CC DIAMETERS.
 CC -1- SUBCELLULAR LOCATION: POSSIBLY SECRETED AND ASSOCIATES WITH THE
 CC EXTRACELLULAR MATRIX.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE DERMATOME. THE EXPRESSION

CC DOMAIN IS MUTUALLY EXCLUSIVE TO THE OTHER WNT GENES.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING EMBRYOGENESIS.
 CC -1- SIMILARITY: BELONGS TO THE WNT FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: D31901; BA06699.1; .
 CC PROSITE: PS00246; WNT1; 1.
 CC PRAM: PF00110; Wnt; 1.
 CC Developmental protein: Glycoprotein; Signal.
 CC SIGNAL 1 24
 FT CHAIN 25 354 WNT-11 PROTEIN.
 FT CAROHD 40 40 POTENTIAL.
 FT CAROHD 50 90 POTENTIAL.
 FT CAROHD 300 300 POTENTIAL.
 FT CAROHD 304 304 POTENTIAL.
 CC SEQUENCE 354 AA: 39507 MW: 14039075 CRC32;
 SO
 WN11.CHICK Length: 354 February 14, 2000 08:03 Type: P Check: 4630 ..
 1 MKSPQFELA AFLSLIQTG ICYGIKIAL SKTPSSIALN QTOCHQLEG
 51 LVVSQVQLCR SNLEMQTII QAAREVIKTC RTESDMRWN CSSIELAPNY
 101 LIDLERTRE SAFYALSA AISHITARAC TTGDIPGSC GPPIGETPGP
 151 GYRWGCAQD LNYGLNGSK FSDAPMKMK SSGQAKRLMH LHNSEVRQV
 201 LKASLEMKR CHVSGGSCI KTCWKGLOEL RDIALDLNK YLSATKVYHR
 251 PMGTRELYVP KDIDIRPYKE TELIYQSSP DECMKNKAVG SHGTQDRQN
 301 KISNGSDCD LMCGRGYNP YMDKYVERCH CKYHCCIVT CKKCRIVER
 351 YVCK
 11AA-SEQUENCE 1.0
 ID WN11.CORJA STANDARD: PRT: 354 AA.
 AC P51891;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE WNT-11 PROTEIN PRECURSOR.
 GN WNT11 OR WNT-11.
 OS Coturnix coturnix japonica (Japanese quail).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
 OC Neognathae; Galliformes; Phasianidae; Phasianinae; Coturnix.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 97176987.
 RA EISENBERG C.A., GOURDIE R.G., EISENBERG L.M.;
 RT "Wnt-11 is expressed in early avian mesoderm and required for the
 RT differentiation of the quail mesoderm cell line QCE-6.";
 RT Development 124:2523-2536(1997).
 CC -1- FUNCTION: MAY PLAY A ROLE IN THE FORMATION OF DERMAL STRUCTURE,
 CC BOTH LIMB AND FEATHER BUDS. IS LIKELY TO SIGNAL OVER ONLY FEW CELL
 CC DIAMETERS (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: POSSIBLY SECRETED AND ASSOCIATES WITH THE
 CC EXTRACELLULAR MATRIX.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING EMBRYOGENESIS.
 CC -1- SIMILARITY: BELONGS TO THE WNT FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: X97549; CAA65151.1; -
 DR PROSITE: PS00246; WNT1.1.
 CC Developmental protein; glycoprotein; signal.
 DR PAM: PF00110; wnt: 1.
 CC Developmental protein; glycoprotein; signal.
 FT SIGNAL 1 24
 FT CHAIN 25 354
 FT CARBOHYD 40 40
 FT CARBOHYD 50 50
 FT CARBOHYD 300 300
 FT CARBOHYD 304 304
 FT SEQUENCE 354 AA; 39457 MW; 3FC9C358 CRC32;
 WNT1.COTVA Length: 354 February 14, 2000 08:03 Type: P Check: 4030

1 MRSPQFLA AFLSLIQTG ICYGIKIAL SKTPSALALN QTOHCKOLEG
 51 LVSAQVQLCR SNLEIMQTI QAAREVIKTC RTFSDMRWN CSSIELAPNY
 101 LLDLEGTRE SAFYALSA AISHIARAC TTGDLPGCSC GPVGEPPGP
 151 GNRWGRCADN LSYGLMGAK FSDAPMKVK TGSQANKLKH LNSEVGRQA
 201 LRASLEMKCK CHGVSGSCI RTCKWGLQEL QDVAADLTKR YLSATKVVHR
 251 PMGTRKHLVP KDIDIRPKVD WEVYLQSSP DFCMKNEKVG SHGTQDRCCN
 301 KTSNGSDSCD IMCGRGINP YTDNRVERCH CKYHMCCTVT CRRCRTVER
 351 YVCK

!!AA:SEQUENCE 1.0
 ID WNT1_HUMAN STANDARD; PRT; 354 AA.
 AC 096014;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DE 15-JUL-1999 (Rel. 38, Last annotation update)
 DE WNT-11 PROTEIN PRECURSOR.
 GN WNT1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 98434463.
 RA IAKO M., STRACHAN T., BULLEN P., WILSON D.I., ROBSON S.C., LINDSAY S.;
 RT "Isolation, characterisation and embryonic expression of WNT11, a gene
 RT which maps to 11q13.5 and has possible roles in the development of
 RT skeleton, kidney and lung."
 RL Gene 219:101-110(1998).
 CC -1- FUNCTION: PROBABLE DEVELOPMENTAL PROTEIN. MAY BE A SIGNALING
 CC MOLECULE WHICH AFFECTS THE DEVELOPMENT OF DISCRETE REGIONS OF
 CC TISSUES. IS LIKELY TO SIGNAL OVER ONLY FEW CELL DIAMETERS.
 CC -1- SUBCELLULAR LOCATION: POSSIBLY SECRETED AND ASSOCIATES WITH THE
 CC EXTRACELLULAR MATRIX.
 CC -1- SIMILARITY: BELONGS TO THE WNT FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL: Y13843; CAA74159.1; -
 DR EMBL: Y13844; CAA74159.1; JOINED.
 DR EMBL: Y13845; CAA74159.1; JOINED.
 DR EMBL: Y13846; CAA74159.1; JOINED.
 DR EMBL: Y13847; CAA74159.1; JOINED.

EMBL: Y12692; CAA73223.1; -
 DR MIM; 603699; -
 DR PROSITE: PS00246; WNT1.1.
 CC Developmental protein; glycoprotein; signal.
 FT SIGNAL 1 24
 FT CHAIN 25 354
 FT CARBOHYD 40 40
 FT CARBOHYD 90 90
 FT CARBOHYD 160 160
 FT CARBOHYD 300 300
 FT CARBOHYD 304 304
 FT SEQUENCE 354 AA; 39407 MW; A4F06B92 CRC32;
 WNT1_HUMAN Length: 354 February 14, 2000 08:03 Type: P Check: 3073

1 MRAPVCEA LFLALQTG VCYGIKIAL SKTPSALALN QTOHCKOLEG
 51 LVSAQVQLCR SNLEIMQTI HAAREVMKAC RRAADMRWN CSSIELAPNY
 101 LLDLEGTRE SAFYALSA TISHIARAC TSGDLPGCSC GPVGEPPGP
 151 GNRWGRCADN LSYGLMGAK FSDAPMKVK TGSQANKLKH LNSEVGRQA
 201 LRASLEMKCK CHGVSGSCI RTCKWGLQEL QDVAADLTKR YLSATKVVHR
 251 PMGTRKHLVP KDIDIRPKVD WEVYLQSSP DFCMKNEKVG SHGTQDRCCN
 301 KTSNGSDSCD IMCGRGINP YTDNRVERCH CKYHMCCTVT CRRCRTVER
 351 YVCK

!!AA:SEQUENCE 1.0
 ID WNT1_MOUSE STANDARD; PRT; 354 AA.
 AC P4615;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE WNT-11 PROTEIN PRECURSOR.
 GN WNT11 OR WNT-11.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-SWISS.
 RX MEDLINE: 96028534.
 RA CHRISTIANSEN J.H., DENNIS C.L., WICKING C.A., MONKLEY S.J.,
 RA WILKINSON D.G., WAINWRIGHT B.J.;
 RT "Murine Wnt-11 and Wnt-12 have temporally and spatially restricted
 RT expression patterns during embryonic development."
 RL Mech. Dev. 51:341-350(1995).
 CC -1- FUNCTION: PROBABLE DEVELOPMENTAL PROTEIN. MAY BE A SIGNALING
 CC MOLECULE WHICH AFFECTS THE DEVELOPMENT OF DISCRETE REGIONS OF
 CC TISSUES. IS LIKELY TO SIGNAL OVER ONLY FEW CELL DIAMETERS.
 CC -1- SUBCELLULAR LOCATION: POSSIBLY SECRETED AND ASSOCIATES WITH THE
 CC EXTRACELLULAR MATRIX.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING EMBRYOGENESIS.
 CC -1- SIMILARITY: BELONGS TO THE WNT FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL: X70800; CAA50070.1; -
 DR MED: MG1:101948; WNT11.
 DR PROSITE: PS00246; WNT1.1.
 DR PAM: PF00110; wnt: 1.
 CC Developmental protein; glycoprotein; signal.
 FT SIGNAL 1 24

WNT1_MOUSE Length: 354 February 14, 2000 08:03 Type: P Check: 1689

1 MRAPQCEA LFLALALHTG VCGIKMLAL SKTPALALN QTHCKQJEG
51 LVSNOVOLCR SNLEKRTIV HAAGAKKAC RRAFADMRN CSSILAFNY
101 LLDIERSTRE SAFYALSA TISHTIARAC TSGDLPCCSC GPVGEPPGP
151 GNRMGACADN LSYGLMGAK FSDAPMYK TGSQANKLMR LHNSEVGROA
201 LRASLETRCK CHGVSGCSI RTCKRGLOEL QDVADLKT R YLSATRYVR
251 PMGRKHLVP KDLDIRPVD SELVYLOSSP DFCMKNEKYG SHGTODROCN
301 KTSNGSDSCD LMCGRGYNP YTDYVERCH CKYHMCQYVT CERCETVER
351 YVCK

11AA_SEQUENCE 1.0 STANDARD: PRT: 389 AA.

ID WNT4_DROME
AC P40589;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE WNT-4 PROTEIN PRECURSOR (DMWT-4).
OS WNT4 OR WNT-4.
GN Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Phryganea; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 95171909.
RA GRABA Y., GIESLER K., ARAGNO D., LAURENTI P., MARIOL M.-C.,
RA BRENGER H., SAGNIER T., PRADÉ J.;
RT "Wnt-4, a novel Drosophila wnt gene acts downstream of homeotic
RT complex genes in the visceral mesoderm.";
RL Development 121:209-218(1995).
CC -1- FUNCTION: ACTS DOWNSTREAM OF HOMEOTIC COMPLEX GENES IN THE
CC VISCERAL MESODERM AND IS REQUIRED FOR EMBRYONIC SEGMENTATION.
CC -1- SUBCELLULAR LOCATION: SECRETED (PROBABLY).
CC -1- SIMILARITY: BELONGS TO THE WNT FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L25316; AAA67470.1;
DR FLYBASE: FBgn0010453; Wnt4.
DR PROSITE: PS00246; wnt; 1.
DR PFAM: PF00110; wnt; 1.
KW Developmental protein; Glycoprotein; Signal.
FT SIGNAL 1 27
FT CHAIN 28 389 WNT-4 PROTEIN.
FT CARBOHYD 134 134 POTENTIAL.
FT CARBOHYD 269 269 POTENTIAL.
SQ SEQUENCE 389 AA: 43245 MW: 803759E3 CRC32;

WNT4_DROME Length: 389 February 14, 2000 08:03 Type: P Check: 8394
1 MGSPRHGHAL AGLAKIGLIV AGGQGLPGNL GYGSTMLNGS GVGGAAGMGL

51 RIGSNTNMD MOGLYNEHE ISEHTWAVE TSGVGSGPC RYMPATRRO
101 HOCREKTGIP GLSEARRLA THCEOFRY DRWNCSIBTR GKNIIFKLY
151 KETAFVHALT AAAMTHSIAR ACAEGMTKC SCGPARKHRE ADFPOMGCN
201 DNLKHKRYVT RSFLDRGSD GDEVEILRN DSEVIEAVS SOMMDCKCH
251 GVGSGCSMT CKMKADPNA TATLLROKN EALAPAPOR SMQVSSRM
301 KRPQRKRP QOSQYTIY LSPSPYCAN TKDRCLRPD NCGTLCCGRG
351 YTVQVQYVE KRCRFNNGR CQCLIDYCO RYENRYFCK

11AA_SEQUENCE 1.0 STANDARD: PRT: 298 AA.

ID Y021_NPYOP
AC 010280;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 34.7 KD PROTEIN (ORF19).
OS Oryza pseudotsugata multicausid polyhedrosis virus (OPNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 97271300.
RA AHRENS C.H., RUSSELL R.R., FUNK C.J., EVANS J., HARWOOD S.,
RA ROHRMAN G.F.;
RT "The sequence of the Oryza pseudotsugata multicausid polyhedrosis virus genome.";
RL Virology 229:381-396(1997).
CC -1- SIMILARITY: NO CORRESPONDING ORF IN ACNPV.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U75930; AAC59018.1;
RX Hypothetical protein.
KW SEQUENCE 298 AA: 34666 MW: 9D51A0D9 CRC32;

Y021_NPYOP Length: 298 February 14, 2000 08:03 Type: P Check: 562

1 MIVQVNYPIQ IYLAHALFGL CSFAVFAFM AVTARYAFL LELEDSHNSI
51 INSHLAFL LGPYLATIT WAMYKMLCY KLELRSNFY MKIVVALAHL
101 MASCHULAV VQPOIHNG HAPYLDALR HHDRSLCMS GVVQVEYEVH
151 DANAIRTDLN CVIYNFMKK CVGCRMEVRA DEPTVFNOQ GALIMLALLA
201 IYHCHNMATV OOKETRRKPN RARNTITL METEVEYDTA EEEHESNM
251 SWLSARRAR NRPNYPPFC ROTGOTONVS RLIFOKKRA ROYEPFAE

11AA_SEQUENCE 1.0 STANDARD: PRT: 153 AA.
ID Y040_HUMAN
AC 015053;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL PROTEIN KIAA0040.
GN KIAA0040.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]

RP SEQUENCE FROM N.A.
RC TISSUE-BONE MARROW;
RX MEDLINE: 96051387;
RA NOMURA N., MIYAJIMA N., SAZUKA T., TANAKA A., KAWABAYASHI Y.,
RA SATO S., NAGASE T., SEKI N., ISHIKAWA K.-I., TABATA S.;
RT Prediction of the coding sequences of unidentified human genes. I.
RT The coding sequences of 40 new genes (K1A0001-K1A0040) deduced by
RT analysis of randomly sampled cDNA clones from human immature myeloid
RT cell line KG-1.
RL DNA Res. 1:27-35(1994).
RP SEQUENCE FROM N.A.
RA BIRD C.;
RL Submitted (JAN-1998) to the EMBL/Genbank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DR EMBL: D25539; BAA05022.1;
DR EMBL: 299715; CAB16870.1;
KW Hypothetical protein.
SQ SEQUENCE 153 AA; 17035 MW; 3D601F8 CRC32;
Y040_HUMAN Length: 153 February 14, 2000 08:03 Type: P Check: 4815 ..
1 MRYVHVRYT TOPRNKPKTK CPSCGSGSGP RCGFLDTYLA AMCFIMALL
51 ADGPMPTCL WATPHAKKE HLSIHLMNP KCVNHHVHT HTNSGATVIG
101 KYLLIKMSL AMFYQGSTL STYTKMSHGK ALPDSDTYIO FPNQCPHTP
151 SIP
ID NA_SEQUENCE 1.0
AC P47451; STANDARD: PRT; 308 AA.
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL PROTEIN MG209.
GN MG209.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
NC [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE: 96026346;
RA FRASER C.M., GOCARINE J.D., WHITE O., ADAMS M.D., CLAYTON R.A.,
RA FLEISCHMANN R.D., BOLT C.J., KERAVAGE A.R., SUTTON G., KELLEY J.M.,
RA FRITZMAN J.L., WEIDMAN J.F., SMALL K.V., SANDISKI M., FUHRMANN J.L.,
RA NGUYEN D.T., UTTERBACK T.R., SAUDEK D.M., PHILLIPS C.A., MERRICK J.M.,
RA TOMB J.F., DOUGHERTY B.A., BOLT K.F., HU P.-C., LUCIER T.S.,
RA PETERSON S.N., SMITH H.O., HUTCHISON C.A. III, VENTER J.C.;
RT "The minimal gene complement of Mycoplasma genitalium."
RT Science 270:397-403(1995).
RL [2]
RP SEQUENCE OF 125-243 FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE: 94075230.
RA PETERSON S.N., HU P.-C., BOLT K.F., HUTCHISON C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing."
RT J. Bacteriol. 175:7918-7930(1993).
RL [1- SIMILARITY: BELONGS TO THE RLU FAMILY OF PSEUDOURIDINE SYNTHASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DR EMBL: U39700; AAC71427.1;
DR EMBL: U02214; AAD12509.1;
DR TIGR: MG209;
DR PROSITE: PS01129; PST_RLU; 1.
DR PFAM: PF00849; YABOY_1.
DR PFAM: PF01479; S4; 1.
KW Hypothetical protein.
SQ SEQUENCE 308 AA; 34996 MW; DF6BFECF CRC32;
Y209_MYCGE Length: 308 February 14, 2000 08:03 Type: P Check: 590 ..

1 MKCGFVYTT KRUDSLASL LNSRKRVK LNMGOIKVN EKLTFKNSLI
51 VAKDVIVKE IHDETSDFI TSVEPYNKL EYLFEDKDIM VINKPSGLT
101 HPTFENEKAS LLAACIFHNN KNPVLYVRL DRDTSGAIV CKNQASLNL
151 ONOLNRLK RYVALVHP FNALGSINA PLARVNNKV MEKIQATA
201 KQATKRFVI NQNEKAALS LELLGRHQ IRVHLKFIQH PYNDPLVGI
251 KSEKDSYGO FLHNRICFI HPTLNKPMDF HAPLEKPFST KIKSLNLSLT
301 DPLHYLFK

ID NA_SEQUENCE 1.0
AC P75485; STANDARD: PRT; 309 AA.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL PROTEIN MG209 HOMOLOG.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
NC [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 29342 / M129;
RX MEDLINE: 97105885.
RA HIMMELREICH R., HILBERT H., PLAGENS H., PIRKL E., LI B.-C.,
RA HERRMANN R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae."
RT Nucleic Acids Res. 24:4420-4449(1996).
RL [1- SIMILARITY: BELONGS TO THE RLU FAMILY OF PSEUDOURIDINE SYNTHASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DR EMBL: AE000053; AAB96191.1;
DR PROSITE: PS01129; PST_RLU; 1.
DR PFAM: PF00849; YABOY_1.
DR PFAM: PF01479; S4; 1.
KW Hypothetical protein.
SQ SEQUENCE 309 AA; 35228 MW; 3375E3F5 CRC32;
Y209_MYCPN Length: 309 February 14, 2000 08:03 Type: P Check: 6533 ..

1 MEQTSVYTA QRLDTFLAL LNLNRVYKAK LIYDGLSVN GKKIKRNMWL
51 VQPEBRVAVN WSELEFEKVP VEQPYDPEPL DILEDEQIM VYNRNGLS

```

101 HPTSFNESES LIGALFHCN HOPFLVHRL DRDTSQVIML AKNOSLHLT
151 OKOLOORVMK RYLLALVHP LDISGTSISA PLERVGNKY MKVGNSSNK
201 AKNAFTFTV LNONEKALI KCELLTGTH QIRVHLOFIK HPVNDPLYG
251 LKSOATREYQ OYLHAQOISF IMPTLNKEMG FEAQDKTFS DKLNLNLKI
301 ANSLYALFO

11AA_SEQUENCE 1.0
ID Y280_MYCGE STANDARD; PRT; 265 AA.
AC B47532;
DC 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL PROTEIN MG280.
GN MG280.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE; 96026346.
RA FRASER C.M., GOCAYNE J.D., WHITE O., ADAMS M.D., CLAYTON R.A.,
RA FEICHSCHMANN R.D., BULL C.J., KERLAVAGE A.R., SUTTON G., KELLEY J.M.,
RA FRITCHMAN J.L., WEIDMAN J.F., SMALL K.V., SANDUSKY M., FORSMANN J.L.,
RA NGUYEN D.T., UTTERBACK T.R., SAUNDER D.M., PHILLIPS C.A., MERRICK J.M.,
RA TOMB J.F., DOUGHERTY B.A., BOTT K.F., HU P.-C., LUCIER T.S.,
RA PETERSON S.N., SMITH H.O., HUTCHISON C.A. III, VENTER J.C.;
RT "The minimal gene complement of Mycoplasma genitalium."
RL Science 270:397-403(1995).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; U9707; AAC71502.1; AUT_INIT.
CC TIGR; MG280;
CC KW Hypothetical protein; Transmembrane.
CC FT TRANSMEM 12 32 POTENTIAL.
CC FT TRANSMEM 201 221 POTENTIAL.
CC SQ SEQUENCE 265 AA; 29566 MW; 5A64E0FF CRC32;

Y280_MYCGE Length: 265 February 14, 2000 08:03 Type: P Check: 9511 ..

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NICHOLS;
RX MEDLINE; 96332770.
RA FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.,
RA DODSON R., GWINN M., HICKEY E.K., CLAYTON R., KETCHUM K.A.,
RA SODERGREN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,
RA KHALAK H., RICHARDSON D., HOWELL J.R., CHIDAMBARAM M., UTTERBACK T.,
RA MCDONALD L., ARTACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,
RA HATCH B., HORST K., ROBERTS K., WATTHEY L., WEIDMAN J., SMITH H.O.,
RA VENTER J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT Spirochete."
RL Science 281:375-388(1998).
CC -1- SIMILARITY: BELONGS TO THE UPF0072 (MESJ/YCF62) FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; AE001216; AAC6358.1;
CC KW Hypothetical protein.
CC SQ SEQUENCE 477 AA; 52605 MW; 6763C3F8 CRC32;

Y373_TREPA Length: 477 February 14, 2000 08:03 Type: P Check: 4876 ..

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AE000023; AAB95867.1; -
DR PFAM: PF00271; Helicase_C.1.
KW Hypothetical protein; Helicase; ATP-binding; RNA-binding.
FT NP_BIND 47 54 ATP (POTENTIAL).
FT SITE 157 157 DEV BOX.
SQ SEQUENCE 450 AA; 51244 MW; 4D9D5B21 CRC32;

Y425_MYCPN Length: 450 February 14, 2000 08:03 Type: P Check: 2944 ..

1 MDSFNEELGV SPALIALTKD NNINOPTTIO QLAIPQFLQH QNLIHSPTG
51 TGTAVFGIP VETLTKKPS KGTDTLVYA PTELEAQIK TPEINAKHT
101 HKAVSLIG IPWQOLKOL ENQPEIVGQ MGRVADLER GVAKFPLEH
151 LIIDEVDLML DRGFRKFLD LLSRIKFEQ IAVYSASVNE ETETAKQIT
201 KNGIFLAPE LKONAPEDN KLIDQFVCL FSNRRKQALY SLVSQFRAS
251 IIVFCOTKTL VDELICFLRK NDVKTYPHLG DKAQIRERN LKLFANTAP
301 IYLVTTDLIG RGIHVEGDM VVNSACVNF ETYLHMGRT GRNNHGSQI
351 TCTSHKROA FLKLEQVND KRISPLRPR LRLPIKKT QPKKGLSLQ
401 SVQKTYVNR SNGTFKRVPL AGDLFSRNR QPERDMQNK LHSDMQSNM

11AA_SEQUENCE 1.0
ID Y464_MYCPN STANDARD: PRT; 385 AA.

AC P47702;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHETICAL PROTEIN MG464.

GN MG464.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE: 96028346;
RA FRASER C.M., GOCAYNE J.D., WHITE O., ADAMS M.D., CLAYTON R.A.,
RA FLEISCHMANN R.D., BUTT C.J., KERAVAGE A.R., SUTTON G., KELLEY J.M.,
RA FRITZMAN J.L., WEIDMAN J.F., SMALL K.V., SANDUSKY M., FUHMANN J.L.,
RA NGUYEN D.T., UTTERBACK T.R., SAUDEK D.M., PHILLIPS C., MERRICK J.M.,
RA TOMB J.F., DOUGHERTY B.A., BOTT K.E., HU P.-C., LUCIER T.S.,
RA PETERSON S.N., SMITH H.O., HUTCHISON C.A. III, VENTER J.C.;
RL "The minimal gene complement of Mycoplasma genitalium";
Science 270:397-403(1995).

RT -1 SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1 SIMILARITY: TO A M.CAPRICOLUM HOMOLOG.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: U39728; AAC72484.1; -
DR EMBL: U39733; AAB01654.1; -
DR TIGR: MG464; -
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 39 59 POTENTIAL.

FT TRANSMEM 150 170 POTENTIAL.
FT TRANSMEM 222 242 POTENTIAL.
FT TRANSMEM 270 290 POTENTIAL.
FT TRANSMEM 325 345 POTENTIAL.
SQ SEQUENCE 385 AA; 44207 MW; D341F781 CRC32;

Y464_MYCPN Length: 385 February 14, 2000 08:03 Type: P Check: 2836 ..

1 MPILAKTINK EIKTFENPEW SAAVNEKNN WNKRFKFSAI FIVIKYFIE
51 IFILIVLNG CQTQLAQPT GTNOVLGSL EIGYKFGTG DYRDLSLNN
101 GPFYFTSDY TLAYGPFYGM FVWPAQIVL PIMYATRVPL GSGVELGFNM
151 ILSLIVLL VRLITIVTL NSTLAEKMN EVOGLAEIN AKYGAIDLQ
201 SKNRQLEIM SLKKNHKS SAAVQVFT LPLFLIYRI VTLRPIKAI
251 ILFNFWDLK VPLTEFSNF TTGMPPIIF LVILPVOFL SQKLPOVWAS
301 KRNENARAS OKSIOQLNT KMOLEFVY FAITAFSNA GGVYWEUNA
351 LFTLOSYLET HVEIVYRREK RKQVSKLDL ILERE

11AA_SEQUENCE 1.0
ID Y464_MYCPN STANDARD: PRT; 385 AA.

AC P75112;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL PROTEIN MG464 HOMOLOG.

OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 29342 / M129;
RX MEDLINE: 97105885;
RA HIMELREICH R., HILBERT H., PLAGENS H., PIKEL E., LI B.-C.,
RA HERRMANN R.;

RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -1 SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1 SIMILARITY: TO A M.CAPRICOLUM HOMOLOG.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AE000017; AAB95810.1; -
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 39 59 POTENTIAL.
FT TRANSMEM 150 170 POTENTIAL.
FT TRANSMEM 222 242 POTENTIAL.
FT TRANSMEM 270 290 POTENTIAL.
FT TRANSMEM 325 345 POTENTIAL.
SQ SEQUENCE 385 AA; 44239 MW; F31AE9B6 CRC32;

Y464_MYCPN Length: 385 February 14, 2000 08:03 Type: P Check: 5866 ..

1 MPILINKKH ELKTFENPEW SAAVNEKNG LKNLKRKMTI IFKYLKVAIF
51 IFILIVLNG CQTQLAQPT GTNOVLGSL EIGYKFGTG DYRDLSLNN
101 VGPFYFTSDY TLAYGPFYGM FVWPAQIVL PIMYATRVPL GSGVELGFNM
151 ILSLIVLL VRLITIVTL NSTLAEKMN EVOGLAEIN AKYGAIDLQ


```

201 SKRNMOMEI SLKKNHKS SASFOVEVT LPILITRI VTLIRIKAI
251 ILFNMWLSK VPLETFSNF TSGWTFITF LIIVPOFI SOKLPITAMS
301 KRNENAKAS OKSIEOLNKT KRMOLIFEV FAVITASAA GVCVYFWLAN
351 LFTLIOSYLT HLFYKRRTR RLTYSKLEQ MLERE

!!AA_SEQUENCE 1.0
ID Y4EK_RHISN STANDARD; PRT: 248 AA.
AC P55434;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE PUTATIVE SHORT-CHAIN TYPE DEHYDROGENASE/REDUCTASE Y4EK (EC 1.-.-.-).
GN Y4EK.
OS Rhizobium sp. (strain NGR234).
OC Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
RN [1]
RP SEQUENCE FROM N.A.
RX FREIBERG C.A., FELLAY R., BAIRNOCH A., BROUGHTON W.J., ROSENTHAL A.,
RA PERRET X.: "Molecular basis of symbiosis between Rhizobium and legumes."
RT Nature 387:394-401(1997).
RL NATURE 387:394-401(1997).
CC -1 SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC FAMILY (SDR). STRONG, TO BACTERIAL 'YDFG'.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE000071; AAB91654.1; -.
CC DR HSSP: P14061; 1FDV.
CC DR PROSITE: PS00061; ADH_SHORT; 1.
CC PFAM: PF00106; adh_short; 1.
CC KW Hypothetical protein; Oxidoreductase; Plasmid.
CC NP_BIND 8 32 NAD OR NADP (BY SIMILARITY).
CC FT ACT_SITE 147 147 BY SIMILARITY.
CC SO SEQUENCE 248 AA; 26679 MW; ATC5BF45 CRC32;

Y4EK_RHISN Length: 248 February 14, 2000 08:03 Type: P Check: 5477 ...
1 MHPMREYA LYGASSGIG KATALEASA GLRYALGRD RAALDELHST
51 AGIVPVEPL SDVEYVYGI AGEKIDYLVN NAGLITASAS LVNLSDDID
101 AMIDINIRSV EKLTRHYLKO MIERRRGHIF FTGSSGGLAP HPNSSYVGAT
151 KAAVSLFSSA LKCDLIGLPI RTVELPFGVY ETNLRYTALG KEGAKKKLYD
201 DNEAIOPHMA ARLITALEL PDEVYTRLE VMPTQGVYVG AQMSKILSR

!!AA_SEQUENCE 1.0
ID Y4TG_RHISN STANDARD; PRT: 231 AA.
AC P5561;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE PUTATIVE 2-ISOPROPYLMALATE/HOMOCITRATE SYNTHASE Y4TG.
GN Y4TG.
OS Rhizobium sp. (strain NGR234).
OC Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE: 97305956.
RA FREIBERG C.A., FELLAY R., BAIRNOCH A., BROUGHTON W.J., ROSENTHAL A.,
RA PERRET X.: "Molecular basis of symbiosis between Rhizobium and legumes."
RT Nature 387:394-401(1997).
RL NATURE 387:394-401(1997).
CC -1 FUNCTION: PROBABLY PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT
CC SYSTEM 14TERG FOR AN AMINO ACID. PROBABLY RESPONSIBLE FOR THE
CC TRANSLLOCATION OF THE SUBSTRATE ACROSS THE MEMBRANE.
CC -1 SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (POTENTIAL).
CC -1 SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-
CC PROTEIN-DEPENDENT TRANSPORT SYSTEMS. BELONGS TO THE HISMO
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE000098; AAB91860.1; -.
CC DR PROSITE: PS00402; BPD_TRANSP_INN_MEMBER; FALSE_NEG.
CC PFAM: PF00528; Bpd.transp; 1.
CC KW Hypothetical protein; Transport; Amino-acid transport; Transmembrane;
CC Inner membrane; Plasmid.
CC KM TRANSMEM 32 52 POTENTIAL.
CC FT TRANSMEM 64 84 POTENTIAL.
CC FT TRANSMEM 86 106 POTENTIAL.
CC FT TRANSMEM 161 181 POTENTIAL.
CC FT TRANSMEM 196 216 POTENTIAL.
CC SQ SEQUENCE 231 AA; 25319 MW; BAC76928 CRC32;

Y4TG_RHISN Length: 231 February 14, 2000 08:03 Type: P Check: 6923 ..
1 MLYGTWDTG NGELAFASIS LPMILGLIT TLQAFLEGF VACVIGWFA
51 VLRGKRTRWV AMPRAVLIEF IRDPLIAQL FELYVYLPXY GIIPPAFLTG
101 ALALGICQSA YISEVYRGCI QAVDHGQREH AKSLDIPPAR TFWYHILPOA
151 IPRVYIPALGN YLVSIKQYV VLSVVTIYEM INAKIIGDQ TENYLVPLSM
201 VGGIYLITLI VASALVRIYD VNLPRGVPL R -

!!AA_SEQUENCE 1.0
ID Y503_METUA STANDARD; PRT: 406 AA.
AC O57926;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE PUTATIVE 2-ISOPROPYLMALATE/HOMOCITRATE SYNTHASE M0503 (EC 4.1.3.-).
GN M0503.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE: 96337999.
RA BUTT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
RA KERAVANGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,
RA OVERBEER R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODER A.,
RA SCOTT J.L., GEORGHEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,
RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
RA CORTON M.D., ROBERTS K.M., HURST M.A., KANE B.P., BORODOVSKY M.,
RA KLINK H.-P., FRASER C.M., SMITH H.O., MOSESE C.R., VENTER J.C.:
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).

```

CC -1- SIMILARITY: BELONGS TO THE ALPHA-IPM SYNTHETASE / HOMOCITRATE
CC SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U67500; AAB98494.1; -
CC TIGR: M0703.1; -
CC PROSITE: PS00816; AIPM_HOMOCIT-SYNTH_1; 1.
CC PFM: PF00682; HMGL-like; 1.
CC KW Hypothetical protein; Lyase.
CC SEQUENCE 406 AA; 45364 MW; 38573C5F CRC32;
Y503_METUA Length: 406 February 14, 2000 08:03 Type: P Check: 6767 ..
1 MTKVLMEMD ELFENSWKAV CPYNPKLDLK DIYIDYTLR DGEQTFGVCF
51 TREOKLEIAR KLDELGLKQI EAGFPVSR EADIVKTIAN EGINADILAL
101 CRALKKIDK AIECDVDGII TFATSPHL KYFNKNSLD EILEMGVEAV
151 EVAKHGFLV AFSADATRT PIEDIKVHK AAEAGADRV HIADTGCAT
201 PQSMFICKI LKENLKKAKH GVCHNDSEF AVINSIGLI GSKANSTIV
251 NGIGERAGNA ALEELIMALT VLYDVLGIN LEVPELCRM VEYSIKRP
301 KKKPIVGEIV FAHESGIHVD AVIENPLTE PFLEKKGK RLNLKSHSG
351 CRAVYIKLKL MGIDYREML CEIVKVKET REESKFTIDE VKREIVEEVL
401 RRRNRN
11AA_SEQUENCE 1.0
ID Y70A_METUA STANDARD; PRT; 102 AA.
AC P8311;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL PROTEIN M0703.1.
GN M0703.1.
OS Methanococcus jannaschii.
OC Archaea: Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE: 96337999;
RA BUTT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
RA KERAVAGE A.R., DOUGHERTY B.A., TOMS J.F., ADAMS M.D., REICH C.I.,
RA OVERBECK R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLUCKER A.,
RA SCOTT J.L., GEOGHAGEN N.S.M., WEIDMAN J.F., FUHRMAN J.L., NGUYEN D.,
RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORDOVSKY M.,
RA KLEMPER H.-P., FRASER C.M., SMITH H.O., MOESE C.R., VENTER J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RT Science 273:1058-1073(1996).
RL -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U67517; AAB98706.1; -
CC DR TIGR: M0703.1; -
CC KW Hypothetical protein; Transmembrane.
FT TRANSMEM 14 34 POTENTIAL.
FT TRANSMEM 35 55 POTENTIAL.
FT TRANSMEM 76 96 POTENTIAL.
FT DOMAIN 76 89 ILE-RICH.
FT SEQUENCE 102 AA; 11685 MW; 2336A893 CRC32;
Y70A_METUA Length: 102 February 14, 2000 08:03 Type: P Check: 4119 ..

1 MGNMNRDK IKSINWNTF IKRITITVGI VISAFAFTIS ILMGMFLIL
51 FLITITFSKT IKRLSKKER SYGLILSLII GSIIITSIIV YSHCYIEFKL
101 LI
11AA_SEQUENCE 1.0
ID Y788_BORBU STANDARD; PRT; 440 AA.
AC O51728;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HYPOTHETICAL PROTEIN B80788.
GN B80788.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 35210 / B31;
RX MEDLINE: 98065643;
RA FRASER C.M., CASDINS S., HUANG W.M., SUTTON G.G., CLAYTON R.A.,
RA LATHIGRA R., WHITE O., KETCHUM K.A., DOBSON R., HICKY E.K., GWINN M.,
RA DOUGHERTY B., TOMS J.F., FLEISCHMANN R.D., RICHARDSON D.,
RA PETERSON J., KERAVAGE A.R., QUACKENBUSH J., SALTBERG S., HANSON M.,
RA VAN VUGT R., PALMER N., ADAMS M.D., GOCAYNE J.D., WEIDMAN J.,
RA UTTERBACK T., WATTHEY L., MCDONALD L., ARTIACH P., BOWMAN C.,
RA GARLAND S., FUJII C., COTTON M.D., HORST K., ROBERTS K., HATCH B.,
RA SMITH H.O., VENTER J.C.;
RA "Genomic sequence of a Lyme disease spirochete, Borrelia
RT burgdorferi.";
RT Nature 390:580-586(1997).
RL -1- SIMILARITY: BELONGS TO THE UPF0072 (MESJ/YCF62) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: A8001177; AAC67121.1; -
CC DR TIGR: B80788; -
CC KW Hypothetical protein.
SO SEQUENCE 440 AA; 51471 MW; D4ABE4A CRC32;
Y788_BORBU Length: 440 February 14, 2000 08:03 Type: P Check: 1619 ..
1 MPELDENIOI KIDKEYKNS LDKNRYIAV SGGADSTALL LNKYYLSNN
51 VIAFYAHLI RSDNEONOEI EHYKGFCDY NIALQIKKCD IDIKSESARL
101 GVAIEELARK FRYIALENAL KENGANYIAL AHENDOIET IIMRFGOSF
151 LDGSGIPSV NRNIIRPLE VSRLEIENEL SLNNGFVVD STNAQNUYLIR
201 NRRVNNLLPA IKVYKGYEK CLKRISEFSK EFADYFGDE FPFVEKGKYY
251 YSDLTATPFD PPKYIVFRLI FRLINSEGLA AKVSKALNE ARVVEINRKK

301 NNVLKTNDF FLEKRRHKNK LIFKRDKEY KPDPILLEVQ KWSLSIGKI
 351 LKYLECNMA SVSRILKCCSY EFRYKPKDR LKAKFFSKFE IRCNPAYIML
 401 LALDNRLIGI IDLNTLNLVW SEKSILKRN ISLIGGLKE
 11AA-SEQUENCE 1.0
 ID Y988_METUA STANDARD; PRT; 170 AA.
 AC P81233;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE HYPOTHEICAL PROTEIN M0795.1.
 GN M0795.1.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
 CC Methanococcus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JAL-1/DSM 2661 / ATCC 43067;
 RX MEDLINE; 96337999.
 RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
 RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
 RA KERLAVAG A.R., DOUGHERTY B.A., TOMB J.F., ADAMS M.D., REICH C.I.,
 RA OVERBEER R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODER A.,
 RA SCOTT J.L., GEOGHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,
 RA UUTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
 RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,
 RA KLEIN H.-P., FRASER C.M., SMITH H.O., MOESE C.R., VENTER J.C.;
 RA "complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 RT Science 273:1058-1073(1996).
 RL Science 273:1058-1073(1996).
 CC -1- SIMILARITY: TO M1249.1, M0210.1 AND M0785.1.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U07524; AAB98801.1; -
 DR Y988_METUA STANDARD; PRT; 329 AA.
 AC O58395;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE HYPOTHEICAL PROTEIN M0798.
 GN M0798.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
 CC Methanococcus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE; 96337999.
 RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
 RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
 RA KERLAVAG A.R., DOUGHERTY B.A., TOMB J.F., ADAMS M.D., REICH C.I.,
 RA OVERBEER R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODER A.,
 RA SCOTT J.L., GEOGHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,
 RA UUTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
 RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,
 RA KLEIN H.-P., FRASER C.M., SMITH H.O., MOESE C.R., VENTER J.C.;
 RA "complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 RT Science 273:1058-1073(1996).
 RL Science 273:1058-1073(1996).
 CC -1- SIMILARITY: SOME. TO M. JANNASCHII M1633.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV
 RX MEDLINE; 96259997.
 RA COLE S.T., BROSCHE R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
 RA GORDON S.V., EIGMEIER K., GAS S., BARRY C.E. III, TERKA R.,
 RA BADDOCK K., BASHAM D., BROWN D., CHILLINGWORTH I., CONNOR R.,
 RA DAVIES R., DEVLIN K., FELTWELL T., GENTLES S., HAMLIN N., HOLROYD S.,
 RA HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOUL S., MURPHY L.,
 RA OLIVER S., OSBORNE J., QUAIL M.A., RAUNDREDDY M.A., ROGERS J.,
 RA RUTHER S., SEEGER K., SKELTON S., SQUARES S., SQUARES R., SULTON J.E.,
 RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 CC -1- SIMILARITY: TO M. TUBERCULOSIS RV2161C AND RV3079C.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; 279700; CAB01997.1; -
 DR Y953_METCTU STANDARD; PRT; 282 AA.
 AC S0913 MW; 6DBA1010 CRC32;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE HYPOTHEICAL PROTEIN M0798.
 GN M0798.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
 CC Methanococcus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE; 96337999.
 RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
 RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
 RA KERLAVAG A.R., DOUGHERTY B.A., TOMB J.F., ADAMS M.D., REICH C.I.,
 RA OVERBEER R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODER A.,
 RA SCOTT J.L., GEOGHAGEN N.S.M., WEIDMAN J.F., FUHRMANN J.L., NGUYEN D.,
 RA UUTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
 RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,
 RA KLEIN H.-P., FRASER C.M., SMITH H.O., MOESE C.R., VENTER J.C.;
 RA "complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 RT Science 273:1058-1073(1996).
 RL Science 273:1058-1073(1996).
 CC -1- SIMILARITY: SOME. TO M. JANNASCHII M1633.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U67541; AAB98990.1; ..
DR TIGR: M09988; ..
DR PIR: P01368; DRH; 1.
KW Hypothetical protein.
SQ SEQUENCE 329 AA; 37316 MW; 8B57A6A CRC32;
Y988_METUA Length: 329 February 14, 2000 08:03 Type: P Check: 9981 ..
1 MELLEYLRKD EYLFCHHNA DPAVGSQVA LKYLASQNP NGFRISADS
51 VSKLSRNILN EIGERVDEI YPKLPETVFI VDTASINOLK VNFDELKERE
101 VILIDHHKKT DLADICKYII IKEDYPTSE IIAEFKELN IFFPKVRIA
151 LKGIYVDRK HLLANSKTF ELISYLKDI SFQKLYLLS QESDVKRTA
201 HAKASRMEI REFDKRLAL SHVSHSASC AKTIVSIAD VAVVAVRKK
251 KEIRVSARC RKHSKYVHL GNLMEKIGE LGSQSGHSE AGGLNAPYDK
301 SSKSEKVIKE VLNLCYKRFV EEYKKAQON
11AA_SEQUENCE 1.0
ID Y988_PSEAE STANDARD; PRT; 122 AA.
AC P21483;
DR 01-MAY-1991 (Rel. 18, Last sequence update)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DE HYPOTHETICAL 13.0 KD PROTEIN IN ALG83 3' REGION.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
CC [1]
RN PSEUDOMONAS
RP SEQUENCE FROM N.A.
RX MEDLINE: 90108714.
RA KATO J., CHU L., KITANO K., DEVAULT J.D., KIMBARA K.,
RA CHAKRABARTY A.M., MISRA T.K.;
RT "Nucleotide sequence of a regulatory region controlling alginate
RT synthesis in Pseudomonas aeruginosa: characterization of the alg83
RT gene.";
RT Gene 84:31-38(1989).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M30145; -; NOT_ANNOTATED_CDS.
DR PIR: J00150; J00150.
KW Hypothetical protein.
SQ SEQUENCE 122 AA; 12967 MW; FEF98CF0 CRC32;
YA32_PSEAE Length: 122 February 14, 2000 08:03 Type: P Check: 6657 ..
1 MALASPAQSR CCAASSARMP ATACCPPTGA SRTCARSRIP WIMRWITGS
51 TATVACARSR RHMPPKRRT TYLPMRACIP NMTPTIYTP MHGRASCWD
101 SASASRMSA GSVISPAVGC CA
11AA_SEQUENCE 1.0
ID Y988_ECOLI STANDARD; PRT; 216 AA.
AC P39220; P15632;
DR 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)

DR 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL 24.8 KD PROTEIN IN DOLA-RLOA INTERGENIC REGION.
GN IAP.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
CC [1]
RN SEQUENCE FROM N.A.
RP STRAIN-K12;
RC MEDLINE: 92334977.
RX YURA T., MORI H., NAGAI H., NAGATA T., ISHIIHAMA A., FUJITA N.,
RA ISONO K., MIZOBUCHI K., NAKATA A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of
RT the 0-2.4 min region.";
RT Nucleic Acids Res. 20:3305-3308(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE: 97426617.
RA BLATTNER F.R., PLUNKETT G., III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER P.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAI B., SHAO Y.,
RT "The complete genome sequence of Escherichia coli K-12.";
RT Science 277:1233-1238(1997).
RN [3]
RP IDENTIFICATION.
RP RUDOLPH K.E.;
RL Unpublished observations (NOV-1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D10483; -; NOT_ANNOTATED_CDS.
DR EMBL: AE000116; AAC73167.1; ..
DR ECOGENE: EG12610; Yabp.
KW Hypothetical protein.
SQ SEQUENCE 216 AA; 24791 MW; 89CF19EB CRC32;
YABP_ECOLI Length: 216 February 14, 2000 08:03 Type: P Check: 98 ..
1 MKVSPGMPV TLNMSKNDI YKWSGDKMD VKKNIIFORLW ETLNHLFWS
51 KOTAYVILF NFVNNOTGNI NASEYFTGAI NENEREKFIN SLELFNKLKT
101 CANKPDELVA KGNMRVAVQT FGDIELSVTF FLEKKNICTQ TLQCHKQGN
151 LGVLDLRAYL PSYDMEDCYL GKRTMGKSD IYERPGWNA NLGVLPRTVL
201 PRIVLRIVL TWTYLP
11AA_SEQUENCE 1.0
ID YBCC_RHOCA STANDARD; PRT; 192 AA.
AC P26166;
DR 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DE HYPOTHETICAL 20.5 KD PROTEIN IN BCFH-CRTU INTERGENIC REGION (ORF192).
OS Rhodospirillum rubrum (Rhodospirillum rubrum capsulatus).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillum group;
OC Rhodospirillum.
CC [1]
RN SEQUENCE FROM N.A.
RP BURKE D.H., ALBERTI M., ARMSTRONG G.A., HEARST J.E.;
RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

RA KIENK H.-P., FRASER C.M., SMITH H.O., MOESER C.R., VENTER J.C.:
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 RL jannaschii, 1.19 Mb." J. Bacteriol. 182:1058-1073 (1996).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: TO M.JANNASCHII M0678.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U67563; AAB99234.1; -
 DR TIGR: MJ1221; -
 KW Hypothetical protein: Transmembrane.
 FT TRANSMEM 128 148 POTENTIAL.
 FT TRANSMEM 177 197 POTENTIAL.
 FT TRANSMEM 210 230 POTENTIAL.
 FT TRANSMEM 236 256 POTENTIAL.
 FT TRANSMEM 261 281 POTENTIAL.
 SQ SEQUENCE 299 AA; 32538 MW; 85A08F7 CRC32;
 YC21.META Length: 299 February 14, 2000 08:03 Type: P Check: 2920 ..
 1 MMEGERMRYM KIIPKFLN TVEIILKNN AVSIIIEPL KTSIEDGIII
 51 TCNADARDAE KIVLELKLGL GSEKGGSVT IMPANITSC REEGIASTSL
 101 SPLEIYYKAK TMVKITKNVI IKVILASING VIGLIEHNIP TLIGAMITAP
 151 LVDTWGSANI GTVLGKELF IQCKKRELLC SGIVIVCAPI PSLFVSKEL
 201 VLQVLSSTSI ILSAIVAIIA GISGSGMSIAS GKEVEIIQVT IDVSILIPAL
 251 LMGMALATMD LYLIVTEPIL LAIVIVLDV GGVIGLKRYV GKINOKIKY
 11AA_SEQUENCE 1.0
 ID YC38.PORPU STANDARD; PRT; 291 AA.
 AC P51321;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE HYPOTHEICAL 32.4 KD PROTEIN YCF38 (ORF291).
 GN YCF38.
 OS Porphyra purpurea.
 OC Chloroplast.
 OG Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AVONPORT;
 RA REITH M.E., MONHOLLAND J.;
 RT "Complete nucleotide sequence of the Porphyra purpurea chloroplast
 RL genome.";
 CC -1- SIMILARITY: BELONGS TO THE YCF38 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U38804; AAC08207.1; -
 DR MENDEL: 10386; PORPU.YC38.1;
 KW Chloroplast; Hypothetical protein
 SQ SEQUENCE 291 AA; 32427 MW; 921437DB CRC32;
 YC38.PORPU Length: 291 February 14, 2000 08:03 Type: P Check: 4681 ..

1 MTFASKKIE LKPIKFTP KYVSEIIOE IEALVORLEL QVWRPAILM
 51 AGIIQPLML VLEGGLEFCA PUNLEITIS YNFFLSGGII VETSTGALN
 101 SGLEIMDRE FGFLNLLLA PLISITIE SSATFMTGIS LIQVIFYIA
 151 SLFGNSPIS SNSTLFFALI VLIVYGYTM LSLATFTLP GHIELLAIL
 201 VVNPFLFSS TALAPIYFAP PWQLIASLN PLSTALEGR YIYSTDMNF
 251 TESVIRKSWG DISLGIISL LIPLDYGAY IVSNILKARL N
 11AA_SEQUENCE 1.0
 ID YCBO.ECOLI STANDARD; PRT; 182 AA.
 AC P75855;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE HYPOTHEICAL FIBRIAL-LIKE PROTEIN IN PEPN-PYRD INTERGENIC REGION
 DE PRECURSOR..
 GN YCBO.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12/MG1655;
 RA MEDLINE: 97426617;
 RA BLATTNER F.R., FLOURENT G., III, BLOCH C.A., PERA N.T., BURLAND V.,
 RA RIEHER M., COLLADO-VIDES J., GLASNER F.D., ROSE D.J.,
 RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN W.A., ROSE D.J.,
 RA MAU B., SHAO Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1234-1238 (1997).
 CC -1- SIMILARITY: BELONGS TO THE FIMB/PAPA FAMILY OF FIMBRIA PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE000196; AAC74024.1; -
 DR ECOGENE: EG13709; YCBO.
 DR PFAM: PF00419; Fimbrin; 1.
 KW Hypothetical protein: Fimbria; Signal.
 FT SIGNAL 1 182 POTENTIAL.
 FT CHAIN 25 182 HYPOTHEICAL FIBRIAL-LIKE PROTEIN YCBO.
 SQ SEQUENCE 182 AA; 19025 MW; CE190E78 CRC32;
 YCBO.ECOLI Length: 182 February 14, 2000 08:03 Type: P Check: 3224 ..
 1 MITMKRSVLT AFTVVCATS SYMAADNVAI TDGVSFTENGK VIAAPCTIVA
 51 AKRDSVTLF DVSAIRKLQIN GOVSGVQIDV PIEIKDCDT VTKNAFTFN
 101 GTADTQITTA FANQASDDA TNVALQWYN DGTATITPT EYGNILLDG
 151 DQITFEKVDY IATGKATSGN VNAVTFNIN YV
 11AA_SEQUENCE 1.0
 ID YC62.ECOLI STANDARD; PRT; 78 AA.
 AC P75991;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE HYPOTHEICAL 8.8 KD PROTEIN IN ICDC-MINE INTERGENIC REGION.
 GN YC62.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

```

OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655:
RX MEDLINE: 9742617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE000215; AAC74248.1; -
CC DR EMBL: EGI4288; YC92.
CC DR Hypothetical protein.
CC KW SEQUENCE 78 AA; 8769 MW; F1F0EC82 CRC32;
CC YC92_ECOLI Length: 78 February 14, 2000 08:03 Type: P Check: 1943
1 MMONSVTLDS AGAITRFPAK ANHTQOETL GEIVTEIILKD GRNLSRSLSC
51 AKLCRLERHA TGEEOEKHYN ALIGLFE
11AA_SEQUENCE 1.0 STANDARD: PRT: 262 AA.
ID YC91_ECOLI
ID P51983; P77675;
DT 01-OCT-1997 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 28.8 KD PROTEIN IN TFX-FNR INTERGENIC REGION.
GN YCJ1.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655:
RX MEDLINE: 9742617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE000215; AAC74248.1; -
CC DR EMBL: EGI4288; YC92.
CC DR Hypothetical protein.
CC KW SEQUENCE 78 AA; 8769 MW; F1F0EC82 CRC32;
CC YC92_ECOLI Length: 78 February 14, 2000 08:03 Type: P Check: 1943
1 MMONSVTLDS AGAITRFPAK ANHTQOETL GEIVTEIILKD GRNLSRSLSC
51 AKLCRLERHA TGEEOEKHYN ALIGLFE

```

```

RL Unpublished observations (MAR-1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE000230; AAC74408.1; -
CC DR EMBL: D90771; CAB20835.1; -
CC DR EMBL: D90770; CAB20825.1; -
CC DR EMBL: U33213; -; NOT_ANNOTATED_CDS.
CC DR EMBL: EGI3906; YCJ1.
CC KW Hypothetical protein.
CC SEQUENCE 262 AA; 28821 MW; CF679910 CRC32;
CC YCJ1_ECOLI Length: 262 February 14, 2000 08:03 Type: P Check: 6021
1 MIIRICITTL PVILPEPETI MYVTPRAER GAPPGETEHY GNSLGAFLI
51 WEPAPASRE SGLIAGTHG DENSSVYLS CALRTILPSL RHHVVLGVN
101 PDGGLGARA NANGVDLNRN FPAANWKEGE TYVRWNSAAE ERDVVILTGD
151 KPGSEPTQA LCOLHRIOP AMVVSFHDPL ACIEDPRHSE LGEWLAQAFE
201 LPVTSVGE TPGSFGSMCA DLNLCTIAE FPPISDSDEAS EYLFAMANL
251 LRMHFKDAIR PS
11AA_SEQUENCE 1.0 STANDARD: PRT: 299 AA.
ID YC91_ECOLI
ID P77333; P16841;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL TRANSCRIPTIONAL REGULATOR IN TFX-FNR INTERGENIC REGION.
GN YCJ2.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655:
RX MEDLINE: 9742617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE000230; AAC74408.1; -
CC DR EMBL: D90771; CAB20835.1; -
CC DR EMBL: D90770; CAB20825.1; -
CC DR EMBL: U33213; -; NOT_ANNOTATED_CDS.
CC DR EMBL: EGI3906; YCJ1.
CC KW Hypothetical protein.
CC SEQUENCE 262 AA; 28821 MW; CF679910 CRC32;
CC YCJ1_ECOLI Length: 262 February 14, 2000 08:03 Type: P Check: 6021
1 MIIRICITTL PVILPEPETI MYVTPRAER GAPPGETEHY GNSLGAFLI
51 WEPAPASRE SGLIAGTHG DENSSVYLS CALRTILPSL RHHVVLGVN
101 PDGGLGARA NANGVDLNRN FPAANWKEGE TYVRWNSAAE ERDVVILTGD
151 KPGSEPTQA LCOLHRIOP AMVVSFHDPL ACIEDPRHSE LGEWLAQAFE
201 LPVTSVGE TPGSFGSMCA DLNLCTIAE FPPISDSDEAS EYLFAMANL
251 LRMHFKDAIR PS

```


CC EMBL: AE000231: AAC74410.1: -
 DR EMBL: D90770: CAB20810.1: -
 DR EMBL: D90771: CAB20837.1: -
 DR EMBL: D90772: CAB20853.1: -
 DR ECOCENE: EG13923: YC32: -
 DR PROSITE: PS00044: HTHLTSR_FAMILY: 1.
 DR PRAM: PF00126: HTHL1: 1.
 KW Hypothetical protein, transcription regulation, DNA-binding.
 FT DNA_BIND 21 41
 SQ SEQUENCE 299 AA: 33481 MW: 48933751 CRC32:

YC32_ECOLI Length: 299 February 14, 2000 08:03 Type: P Check: 51 ..

1 MKRETIADIM AFVVAEERS FTBAARLSM AQSLSQIVR RIBERGLRL
 51 LRTTRSVVP TEAGEHLTV LGPMLHDIDS AMASISDLON RPSGITRITT
 101 VEHAAKTILL PAMRTFLSKH PEIDITLID YGLEDVYSER FDAGVRLGGE
 151 MDKMDAIRI GPDIPMAIVG SPDYFSRSV PTVSOLIDH QAINLYLPTS
 201 GTANRWRLIR GGREVRVME GOLLNTIDL IIDAIDGHS LAVLPDYOE
 251 RAIRKELIR VLDKTPDLF GYHLYPHRR HGSASFSLFI DRLKYGAV

11AA_SEQUENCE 1.0
 ID YC05 YEAST STANDARD: PRT: 317 AA.

AC P25616:
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE HYPOTHEICAL 36.3 KD PROTEIN IN POLA-SB1 INTERGENIC REGION.
 GN YC0515C OR YC0515C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
 CC Saccharomycetales; Saccharomycetes.

CC (1)
 RN SEQUENCE FROM N.A.
 RA HAYAT D., JACO C., PEREA J., SHU Y.;
 RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X59720: CAA42332.1: -
 DR PIR: S19425: S19425.
 KW Hypothetical protein.

SO SEQUENCE 317 AA: 36264 MW: 04265155 CRC32:

YC05_YEAST Length: 317 February 14, 2000 08:03 Type: P Check: 1969 ..

1 MKTIISFDF ETTTRVDTIC TAKLPYLLN PRKPEKGF TTYWDGTHK
 51 KYKNTSLP LSSGVPITII SOSNFKLFA DELAYONHR VVELNSVNEI
 101 TKQOIFKIS LDMKTFARD QNHDDCLRD GFKTFCSYV KNFESDFYVL
 151 SINMSKEIH EVIGDRLLN SHIFCNDLKK VSDKCSOSYN GFCDRLTLTG
 201 SDKVILIEI LDRIDSGCNK EGNCSYWI GDSETDLASI LHPSTNGVLL
 251 INPOENPSKF IKITEKIGI PKDKISSFEA DNGPAMLOFC EKEGKGAVL
 301 VKSMDSLKDL IMOVTKM

11AA_SEQUENCE 1.0
 ID YDM1_SCHPO STANDARD: PRT: 131 AA.

AC P87132:
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE HYPOTHEICAL PROTEIN C57A7.01 IN CHROMOSOME I (FRAGMENT).
 GN SPAC57A7.01.
 OS Schizosaccharomyces pombe (Fission yeast).
 CC Eukaryota; Fungi; Ascomycota; Archiaacomycetes;
 CC Schizosaccharomycetales; Schizosaccharomycetales;
 CC Schizosaccharomycetes.
 CC (1)
 RN SEQUENCE FROM N.A.
 RA SKELTON J., CHURCHER C.M., WOOD V., BARRELL B.G., RAJANDREAM M.A.;
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC -1- SIMILARITY: TO M.JANNASCHIT MD0531.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z95396: CAB8759.1: -
 DR Hypothetical protein.
 KW NON_TER 1
 FT SEQUENCE 131 AA: 14774 MW: 1943767C CRC32:

YDM1_SCHPO Length: 131 February 14, 2000 08:03 Type: P Check: 6054 ..

1 DPSARAVKOR WESEOLETLE KIRKYLILKL SKTYLEVEVN IEVHHEKAR
 51 HLIEMIDYI EPSLVWMSR GRSHLKVLL GSFNVLVNR SSVPMVARK
 101 KLRKNRRLG NQSRLLNLS DAIVDEYGR P

11AA_SEQUENCE 1.0
 ID YDSA_SCHPO STANDARD: PRT: 174 AA.
 AC Q14185:
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE HYPOTHEICAL 20.1 KD PROTEIN C4F8.10C IN CHROMOSOME I.
 GN SPAC4F8.10C.
 OS Schizosaccharomyces pombe (Fission yeast).
 CC Eukaryota; Fungi; Ascomycota; Archiaacomycetes;
 CC Schizosaccharomycetales; Schizosaccharomycetales;
 CC Schizosaccharomycetes.
 CC (1)
 RN SEQUENCE FROM N.A.
 RA STRAIN-972;
 RC GENTLES S., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WOOD V.;
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC -1- SIMILARITY: TO YEAST YOR367W AND TO CALPONINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z98530: CAB11057.1: -
 DR Hypothetical protein.
 KW SEQUENCE 174 AA: 20141 MW: 2DB8876B CRC32:

YDSA_SCHPO Length: 174 February 14, 2000 08:03 Type: P Check: 8869 ..

1 MTSQLEKAR EWIEETLHTK LNHQDLDDQ LQSGVILCRI CREALGANIR

```

51 YKSNMPEVQ MENISAFINY AOVVAVPSQ DMFOTSDUFE RRNDQVLR8
101 IHSFRYAK MFPGKVRGLG PKLAERKPRV FSAOQOEFRR EGVNSLOYGS
151 FDMPTGTEK IAFSRDRPT GNMV

!!AA_SEQUENCE 1.0
ID YDRE_SCHPO STANDARD: PRT: 344 AA.
AC 014220:
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHEICAL 59.8 KD PROTEIN CGB12.16 IN CHROMOSOME 1.
GN SPAG6B12.16.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Archaescomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972:
RA GENTILES S., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WOOD V.;
RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL: 298531; CAB1075.1;
KM Hypothetical protein
SQ SEQUENCE 344 AA; 39810 MW; BBE84116 CRC32;

YDRE_SCHPO Length: 344 February 14, 2000 08:03 Type: P Check: 3748 ..

1 MNSNEIENFI YLNTLTISEV PEPNSVFTPT KCFNSKLPH QSDIEPSSA
51 CSLSKRTIID GGADEPYDSS DSCATEFAD FEKNDLDFEA LDGEIINIH
101 VNQTSPTTHH NAEPLDLOI FSSNLHNSN RRQITCNFNM ANDASKENET
151 PYWVLNKNFN PVLTEYTOQ HLVOCKWYLE NHITSRPHF YTKLPDVSIV
201 PNMHPYDDE VTAKAPKDD FYVPRFRGH GISKGLCPI CSHQGEFIWL
251 RTTSAYWTH MNFVGHSHK GRPYQPIEF RTVRLKTRN AIGVNNKYM
301 IEKCHQCNK WIRQGRDQ SVKIPPEFW RHAHCHIT TDLR

!!AA_SEQUENCE 1.0
ID YK41_EAST STANDARD: PRT: 128 AA.
AC 740057:
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHEICAL 14.3 KD PROTEIN IN PTP3-ILV1 INTERGENIC REGION.
GN YK084W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-5288C / AB972:
RA DIETRICH F.S., MULLIGAN J.T., HENNESSEY K.M., ALLEN E., ARAUJO R.,
RA AVILES E., BERNO A., BRENNAN T., CARPENTER J., CHEN E., CHERRY J.M.,
RA CHUNG E., DUNCAN M., GUZMAN E., HARTFELD G., HUNICKER-SMITH S.,
RA HYMAN R., KAYSER A., KOMP C., LASHKARI D., LEW H., LIN D.,
RA MOSEDALE D., NAKAHARA K., NAMATH A., NORGREN R., OEFNER P., OH C.,
RA PETER F.X., ROBERTS D., SEHL P., SCHRAMM S., SMOGREN T., SMITH V.,
RA TAYLOR P., WEI Y., YELTON M., BOTSTEIN D., DAVIS R.W.;

```

```

RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL: U18839; AAB64639.1;
KM Hypothetical protein
SQ SEQUENCE 128 AA; 14313 MW; 564479CB CRC32;

YK41_EAST Length: 128 February 14, 2000 08:03 Type: P Check: 5798 ..

1 MELICITPY HSLDFEFL FCPKRRAR GHRKFLTL YKSNLIRKL
51 LPPSLFTKRV MNPSSHPS PDFPGSSAS PRVKLRPSTL WAPLTVSSD
101 PAASSSTAP VVTDKPVT AVSKRYQP

!!AA_SEQUENCE 1.0
ID YF8_MYCTU STANDARD: PRT: 148 AA.
AC 010772:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE HYPOTHEICAL 16.4 KD PROTEIN RV1558.
GN RV1558 OR MYC48.07C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV.
RA MEDLINE: 9625987.
RA COLE S.T., BROSCHE R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA GORDON S.V., EIGMEIER K., GAS S., BART C.E. II, TENKIN E.,
RA BADOCK K., BASHAM D., BROWN D., CHILINGWORTH T., CONNOR R.,
RA DAVIES R., DEVLIN K., FELTWEIL T., GENTILES S., HAMLIN N., HOLROYD S.,
RA HORSBER T., JAGELS K., KROGH A., MCLEAN J., MOULÉ S., MURPHY L.,
RA OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,
RA RUTTER S., SEBBER K., SKELTON S., SQUARES S., SQUARES R., SULSTON J.E.,
RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
CC -1- SIMILARITY: TO M.TUBERCULOSIS RV1261C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL: Z74020; CA98933.1;
KM Hypothetical protein
SQ SEQUENCE 148 AA; 16347 MW; 7F320812 CRC32;

YF8_MYCTU Length: 148 February 14, 2000 08:03 Type: P Check: 497 ..

1 MPLSEYAPS PLDMREQAD TYKSGTGG TQLOGKPVIL LTVGAKTK
51 LKRTPLMRE HDGOYAIVAS LGAPRPYV YHNVYKNRP ELQDGTVTD
101 YDAREVFGDE KAIMORAVA VMPDYASYOT KTRQIFEVF LPPVRAGG

!!AA_SEQUENCE 1.0
ID YG29_BPSPI STANDARD: PRT: 148 AA.
AC P31653:

```

Mon Feb 14 08:07:43 2000

sp.cat

Page 158

01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE HYPOTHEORETICAL 16.2 KD PROTEIN IN GENE 29.5 REGION.
OS Bacteriophage SP01.
CC Viruses; dsDNA viruses, no RNA stage: Tailed phages: Myoviridae.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:WIT
RX MEDLINE; 92351562.
RA WILHELM K., ROEBER W.;
RT "Deoxyuridylylate-hydroxymethylase of bacteriophage SP01."
RL Virology 189:640-646(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X60728; CAA43135.1; -
DR PIR; S21504; S21504.
KW Hypothetical protein.
SQ SEQUENCE 148 AA; 16183 MW; 7C6906B5 CRC32;
YG29_BPSP1 Length: 148 February 14, 2000 08:03 Type: P Check: 7782 ..

1 MNQGVNLKIG GYVWDPPVK VESKTYGELE VTRKKVYHRM LDIIIPVGLA
51 TKRLTALSPG ALAAGVDSA DKIRGPHDI IDVTALAEF ILMFVATLTK
101 VLIATKKNNA GWERLKNVGT AVAGIALLEPT FFSFLRWVSS IYSSSITF
11AA_SEQUENCE 1.0
ID YGK9_YEAST STANDARD; PRT; 107 AA.
AC P53138;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHEORETICAL 12.4 KD PROTEIN IN TAF60-G4P1 INTERGENIC REGION.
GN YGL109W OR G3065.
OS Saccharomyces cerevisiae (Baker's yeast).
CC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
CC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97197974.
RA PAOLUZI S., MINENKOVA O., CASTAGNOLI L.;
RT "The genes encoding the transcription factor YTAFL160, the G4P1
RT protein and a putative glucose transporter are contained in a 12.3 kb
RT DNA fragment on the left arm of Saccharomyces cerevisiae chromosome
RT VII."
RL Yeast 13:85-91(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X97644; CAA66243.1; -
DR EMBL; 272631; CAA96815.1; -
KW Hypothetical protein.
SQ SEQUENCE 107 AA; 12435 MW; 349DFDE6 CRC32;
YGK9_YEAST Length: 107 February 14, 2000 08:03 Type: P Check: 6897 ..

1 MAAGNPPLADI QYKRYKAR RMEGOKKNSC TIAIDSLQY YCRSLSHKS

51 CFPSPQSHAF SBNPLPSESY EFWHALELAF CLTRPYCTFH SLEISSQOL
101 TLRRPLIG
11AA_SEQUENCE 1.0
ID YGIR_STRCO STANDARD; PRT; 66 AA.
AC P05954;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE HYPOTHEORETICAL PROTEIN IN GLNR 3 REGION (FRAGMENT).
OS Streptomyces coelicolor.
CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
CC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:A3(2);
RX MEDLINE; 93345814.
RA WRAY L.V., JR., FISHER S.H.;
RT "The Streptomyces coelicolor glr gene encodes a protein similar to
RT other bacterial response regulators."
RL Gene 130:145-150(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L02213; AAA02839.1; -
DR PIR; P06044; P06044.
DR HSSP; P15039; IPRV.
KW Hypothetical protein.
FT NON_TER 66
SQ SEQUENCE 66 AA; 7094 MW; 502F47A4 CRC32;
YGIR_STRCO Length: 66 February 14, 2000 08:03 Type: P Check: 9954 ..

1 MAYTRDVA RLAGSTAVV SYVINGNRP VAPATREHVL AAIKEIGYRP
51 DRVAQAMASR RDLIG
11AA_SEQUENCE 1.0
ID YHAL_CRYPA STANDARD; PRT; 319 AA.
AC P10941;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE HYPOTHEORETICAL PROTEIN 1 IN HYPOVIRULENCE-ASSOCIATED DS-RNA GENETIC
DE ELEMENT (CONTAINS P29 PROTEINASE).
OS Cyphonectria parasitica (Chesnut blight fungus) (Endothia
OS parasitica).
CC Eukaryota; Fungi; Ascomycota; Euascomycetes; Pyrenomycetes;
CC Diaporthales; Valsaceae; Cyphonectria.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-EP713;
RX MEDLINE; 89251594.
RA RAE B.P., HILMAN B.I., TARTAGLIA J., NUSS D.L.;
RT "Characterization of double-stranded RNA genetic elements associated
RT with biological control of chestnut blight: organization of terminal
RT domains and identification of gene products."
RL EMO J. 8:657-663(1989).
CC -----
CC MISCELLANEOUS: DOUBLE-STRANDED RNA GENETIC ELEMENTS ARE ASSOCIATED
CC WITH BIOLOGICAL CONTROL OF THE FUNGAL DISEASE CHESTNUT BLIGHT.
CC THIS DS-RNA ARE ASSOCIATED WITH HYPOVIRULENCE. THEY ARE LOCALIZED
CC IN THE CYTOPLASM.
CC -----
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C7.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X14524; CA32666.1; -
 DR PIR: S03833; S03833.
 KW Hypothetical protein; Hydrolase; Thiol protease.
 SQ SEQUENCE 319 AA; 35443 MW; 6309763E CRC32;

YHAI_CRYPA Length: 319 February 14, 2000 08:03 Type: P Check: 3269 ..

1 MAOLIKPSQS LVLSSEVDPT TVDPFVSVRT EEEVPAACIT IMEYDSCGD
 51 VPGFLSHGDL RLRLPPDGVK KQVHFELPT VLKSGSTGTV PEHPAVLAF
 101 IGRPRRCSLE ORTKELDSRF LQVHGGLPA RPSYMLARP REVRGLCSSR
 151 NSGLAOFGOG YCYLSAIVDS ARWRVARTTG WCVRVADYLR LLQWGRRSF
 201 GSFOIERSAV DHYHVYVDA EYQSEODGAL FYQALIGLAE KDPLARIGGR
 251 LNPAAEFAP GSALRVEPYT PQTERRKGST RMTGRDPTIV FRWQGYGHH
 301 QHPCSCGYG VEERFRPS

IIAA_SEQUENCE 1.0 STANDARD; PRT; 302 AA.

ID YHCT_BACSU
 AC P54604;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE HYPOTHELICAL 33.7 KD PROTEIN IN CSPB-GLFP INTERGENIC REGION.

GN YHCT
 OS Bacillus subtilis.
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/staphylococcus group; Bacillus.

RN
 RP SEQUENCE FROM N.A.
 RX STRAIN-168;
 RX MEDLINE; 97124185.

RA NOBACK M.A., TERPSTRA P., HOLSAPPEL S., VENEMA G., BRON S.;
 RT "A 22 kb DNA sequence in the CSPB-glfpkx region at 75 degrees on the
 RT Microbiology 142:3021-3026(1996)."

RL
 CC -1- SIMILARITY: BELONGS TO THE RLU FAMILY OF PSEUDOURIDINE SYNTHASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: X96983; CA65704.1; -
 DR EMBL: Z99108; CAB12749.1; -
 DR SUBILIST; B611598; YHCT.
 DR PROSITE; P501129; P51_RLU; 1.
 DR PIR; P500849; YABO; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 302 AA; 33740 MW; 2EB71306 CRC32;

YHCT_BACSU Length: 302 February 14, 2000 08:03 Type: P Check: 1929 ..

1 MNOKRGLEI LINEKODGW LPSVLTKALK ASKEVIDMM SHOQIVNHE
 51 SVLNMIYVK GDRVFDLQE SEASSVIPEY GELDILFEDN HMLINKPAG
 101 IATHNEDGQ TGLANLIAY HYQINGETCK VRHVRLDOD TSGAIFAAH
 151 RLAAHILDOQ LEKTLKRTY TAAEGKRT KGTINPPIG RDRSHPTRRR

201 VSPGGOTAVT HFKWASNAK ERLSLVELEL ETGRTHQIRV HLASLGHPLT
 251 GDSLYGGGSK LNRQALHAN KQAVHPITD ELIVAEPFP ADMKNLCRTY
 301 FS

IIAA_SEQUENCE 1.0 STANDARD; PRT; 429 AA.

ID YHG3_YEAST
 AC P38756;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)
 DE HYPOTHELICAL 48.9 KD PROTEIN IN RPL14B-GPA1 INTERGENIC REGION.

GN YH003C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
 CC Saccharomycetaceae; Saccharomyces.

RN
 RP SEQUENCE FROM N.A.
 RX STRAIN-S288C / AB972;
 RX MEDLINE; 94378003.

RA JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DOVER J.,
 RA DU Z., FAVELLO A., FULLON L., GATUNG S., GEISEL C., KIRSTEN J.,
 RA KUCABA T., HILLIER L., JIER M., JOHNSON L., LANGSTON Y.,
 RA LATREILLE P., LOUIS E.J., MACRI C., MARDIS E., MENDEZ S., MOUSER L.,
 RA NHAN M., RIFKIN L., RILES L., ST PETER H., TREVASKIS E., VAUGHAN K.,
 RA VIGNATI D., WILCOX L., WOHLDMAN P., WATERSTON R., WILSON R.,
 RA VAUDIN M.;

RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
 RT viii.";
 RL Science 265:2077-2082(1994).

CC -1- SIMILARITY: SPRONG, TO YEAST YKL027M, SOME, TO E. COLI MOLYBDENUM
 CC COFACTOR BIOSYNTHESIS PROTEIN A (MOA).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: U10555; AAB68430.1; -
 DR PIR; S46801; S46801.
 DR PIR; P500899; THLF_family; 1.
 KW Hypothetical protein.

SQ SEQUENCE 429 AA; 48883 MW; 709A28CF CRC32;

YHG3_YEAST Length: 429 February 14, 2000 08:03 Type: P Check: 5569 ..

1 MANTWKLIA TVALISVEST QLAQSVKREY KLSAANKRK TVSPROYD
 51 HLFREGQARN VAFLEGGR KIRQYIVIV GAGEVGSWC TMLIRSCOK
 101 IMIDPENIS IDSLNTHCCA VLSDIGPKV QCLKEHLISKI APMEIRARA
 151 KAWTKENSHD LIFADGSEPT FIVDCLDNLE SKVDLELYAH HNKIDIVISM
 201 GVATKSDPTR VSINDSMTE FDPISRCVRR KLRKRIAGT ISVFSNEML
 251 DPRDDILSP IDCEHRAINA VRDALNHP ELDTMGITG LSIATWILTK
 301 VSGIPKRENE VAKRLKFYDS ILETFQXOMA RLNENKERS LLGLEEYVI
 351 VEENFRGKSP ISGYSTKLAL TKWEANKSIS LINVVLMTKE EOEIHEKIL
 401 LDGEKLTAVY SEEVDFIER LFKEEYIS

IIAA_SEQUENCE 1.0 STANDARD; PRT; 197 AA.
 ID YHGN_ECOLI
 AC P46851;
 DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE HYPOTHELICAL 21.5 KD PROTEIN IN ASD-GNTU INTERGENIC REGION (0197).
 GN YHGN.
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE; 9742617.
 RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
 RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
 RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
 RA MAU B., SHAO Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE UPF0056 (MARC) FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: U18997; AA58932.1;
 DR EMBL: AF000420; AAC76459.1;
 DR EMBL: EC12841; YHGN.
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 4 24 POTENTIAL.
 FT TRANSMEM 45 65 POTENTIAL.
 FT TRANSMEM 72 92 POTENTIAL.
 FT TRANSMEM 106 126 POTENTIAL.
 FT TRANSMEM 139 159 POTENTIAL.
 FT TRANSMEM 174 194 POTENTIAL.
 SQ SEQUENCE 197 AA; 21490 MW; 089FBACD CRC32;
 YHGN_ECOLI Length: 197 February 14, 2000 08:03 Type: P Check: 1678 ..
 1 MNEIISAVAL LILIMDPLGN LPIFMSVLKH TEPKRRRAIM VRELLIALIV
 51 MLVFLFAGER ILAFSLRAE TVSISGILL FLIAIKMIFP SASGNSGSDP
 101 AGEFPFIVLP AIPVAGPTI LATIMLSHQ YPNQMGHLVI ALLANGTIF
 151 VILQSLFL RLIGKGVNA LERLMGLIV MMATOMFLDG IRMMWKG
 11AA_SEQUENCE 1.0 STANDARD; PRT; 164 AA.
 ID Y129_MYCTU
 AC 050604.
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-JUN-1998 (Rel. 38, Last sequence update)
 DT 15-DEC-1998 (Rel. 38, Last annotation update)
 DE HYPOTHELICAL 18.1 KD PROTEIN RV1829.
 GN RV1829 OR MTCT11.14C.
 OS *Mycobacterium tuberculosis*.
 OC Bacteria; Filumetes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RX MEDLINE; 98295987.
 RA COLE S.T., BROSCHE R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
 RA GORDON S.V., EIGLMEIER K., GAS S., BARRY C.E. III, TEKALA F.,
 RA BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
 RA DAVIES R., DEVLIN K., FELTWEIL T., GENTLES S., HAMLIN N., HOLROYD S.,
 RA HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
 RA OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,
 RA RUTTER S., SEGER K., SKELTON S., SQUARES S., SQUARES R., SULSTON J.E.,
 RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;
 RT "Deciphering the biology of *Mycobacterium tuberculosis* from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: Z78020; CAB01481.1;
 DR EMBL: Z78020; CAB01481.1;
 KW Hypothetical protein.
 SQ SEQUENCE 164 AA; 18114 MW; 09A1BF7B CRC32;
 Y129_MYCTU Length: 164 February 14, 2000 08:03 Type: P Check: 7074 ..
 1 MGEVAVGIR VEQPQNPVL LLREANGDRY LPIWIGQSEA AAILDEQGV
 51 EPPRLTHDL IRDLIALGH SLKEVRIYVL QEGTFYADLI FDRNIVSAR
 101 PSDSVATLR VGVPIVEEA VLAQAGLLIP DESDEATRA VREDEVERFX
 151 EFLDSVSPD FKAT
 11AA_SEQUENCE 1.0 STANDARD; PRT; 317 AA.
 ID Y129_MYCTU
 AC 010863.
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-DEC-1998 (Rel. 39, Last annotation update)
 DE HYPOTHELICAL 33.9 KD PROTEIN RV1956.
 GN RV1956 OR MTCT139.23C.
 OS *Mycobacterium tuberculosis*.
 OC Bacteria; Filumetes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RX MEDLINE; 98295987.
 RA COLE S.T., BROSCHE R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
 RA GORDON S.V., EIGLMEIER K., GAS S., BARRY C.E. III, TEKALA F.,
 RA BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
 RA DAVIES R., DEVLIN K., FELTWEIL T., GENTLES S., HAMLIN N., HOLROYD S.,
 RA HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
 RA OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,
 RA RUTTER S., SEGER K., SKELTON S., SQUARES S., SQUARES R., SULSTON J.E.,
 RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;
 RT "Deciphering the biology of *Mycobacterium tuberculosis* from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 CC
 CC -1- SIMILARITY: TO B.SUBTILIS XXIF.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: Z74025; CAA8390.1;
 DR EMBL: Z74025; CAA8390.1;
 KW Hypothetical protein.
 SQ SEQUENCE 317 AA; 33879 MW; AAFB579 CRC32;
 Y129_MYCTU Length: 317 February 14, 2000 08:03 Type: P Check: 1836 ..
 1 MSAQQTNLGI VVGVDGSPCS HTAVENWARD AQMRNVALRV VQVVPVITA
 51 PEGNAFYSR FQDAKREIV EHSIVQAQH QIVEQAHKVA LEASSSRRRA

```

101 QITGEVLHGQ IYPTLANISR QVAMVVLGYR GGGAVAGALL GSVSSLYRH
151 AHGPVAVIPE EPRPARPPHA PYVVGIGISP TSGLAELIIF DEARRRVDL
201 VALHANSDMG PLDFPRLNKA PIEKRNLEDE QEKMLARRIS GMQRIYDYV
251 VHKVYVCDRP APRLELAOT AQLVVGSGH RGGFGMHNG SVSRVAVNSG
301 QAPYVARIP QDPVAPVA

11AA_SEQUENCE 1.0
ID YJUX_ECOLI STANDARD; PRT: 173 AA.
AC P39410;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 18.6 KD PROTEIN IN TRPR-GPMB INTERGENIC REGION (F173).
GN YJUX.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12/MG1655;
RX MEDLINE; 95334362
RA BURLAND, J.D., FLUNKETT G. III, SOFIA H.J., DANIELS D.L.,
RA BLATTNER F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes."
RL Nucleic Acids Res. 23:2105-2119(1995).
RN [2]
RP SEQUENCE OF 31-173 FROM N.A.
RX MEDLINE; 81053831.
RA SINGLETON C.K., ROEDER W.D., BOGOSIAN G., SOMERVILLE R.L., WEITH H.L.;
RT "DNA sequence of the E. coli trpR gene and prediction of the amino
RT acid sequence of the Trp repressor."
RL Nucleic Acids Res. 8:1551-1560(1980).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U1003; AAA97290.1;
DR EMBL; AE000509; AAC77347.1;
DR EMBL; J01715; -; NOT_ANNOTATED.CDS.
DR ECGENE; EG12600; yjux.
KW Hypothetical protein.
SQ SEQUENCE 173 AA; 18570 MW; 2DC3DAB3 CRC32;

YJUX_ECOLI Length: 173 February 14, 2000 08:03 Type: P Check: 3873
1 MLINHOVCA TTNPAKIQAI LQAFHEIRGE GSCHIASVAV ESGVEQDFG
51 SEETRAGARN RVANARRLLP EADEVAIEA GIDGDSFVW VVIENASORG
101 EARSATLPLP AVILEKVRG EALGPVMSRY TGIDIGRKE GAIGVFTAGK
151 LTRASYVHQA VILALSPFN AVY

11AA_SEQUENCE 1.0
ID YJUX_ENTAE STANDARD; PRT: 54 AA.
AC P39430;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 01-FEB-1995 (Rel. 31, Last annotation update)
DE HYPOTHETICAL PROTEIN IN TRPR 3 REGION (FRAGMENT).
GN YJUX.
OS Enterobacter aerogenes (aerobacter aerogenes).
OS Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

```

```

OC Enterobacter.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94268903.
RA ARVIDSON D.N., ARVIDSON C.G., LAMSON C.L., MINER J., ADAMS C.,
RA YONDERMAN P.;
RT "The tyrophan repressor sequence is highly conserved among the
RT Enterobacteriaceae."
RL Nucleic Acids Res. 22:1821-1829(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L26582; AAC36893.1;
KW Hypothetical protein.
KW NON_TER
SQ SEQUENCE 54 AA; 5794 MW; 0E5A3E2A CRC32;

YJUX_ENTAE Length: 54 February 14, 2000 08:03 Type: P Check: 4167
51 AIYR
1 EALGPVMSGH TGIDIGRKE GAIGVFTAGK LTRASYVHQA VILALSPFN

11AA_SEQUENCE 1.0
ID YJUX_ENTAE STANDARD; PRT: 196 AA.
AC P46389;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL 21.6 KD PROTEIN IN ATP12-PL17B INTERGENIC REGION.
GN YJUL78C OR J0490.
OS Saccharomyces cerevisiae (Baker's yeast).
OS Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetales; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RA OBERHAIR B., PIRAVANDI E., RINK M., DOMDEY H.;
RT Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z49453; CAA89473.1;
KW Hypothetical protein; Transmembrane.
KW TRANSMEM 125 145 POTENTIAL.
SQ SEQUENCE 196 AA; 21605 MW; AEA06F63 CRC32;

YJUX_ENTAE Length: 196 February 14, 2000 08:03 Type: P Check: 8234
1 MLCGLTIVIL PGKDAITQI IDEFNKIGNF VEETESALTL TLKGATWGAN
51 SFDKLEFQC NDNKODELT SHTWADKSIO LTIKPSGCL KSRDDKKNG
101 DGDNGKDGDS EGKPKAKAG GTSWFTWLF YALFTLLYL MYVSFLNTRG
151 GSFQDFRAEF IQRSQFLTS LPEFCREVS RIIGSTNOR GGVSAV

11AA_SEQUENCE 1.0
ID YJUX_ECOLI STANDARD; PRT: 79 AA.
AC P35677;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)

```

DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE HYPOTHEETICAL 9.0 KD PROTEIN IN PROA-PERR INTERGENIC REGION.
 GN YKFF.
 OS Escherichia coli.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE: 97426617
 RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
 RA RILEY M., COLLADO-VIDES J., GIANER F.D., RODE C.N., MAHEM G.F.,
 RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOLDEN M.A., ROSE D.J.,
 RA MUR B., SHAO Y.,
 RT "The complete genome sequence of Escherichia coli K-12."
 RL Science 277:1233-1238(1997).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE000133; AAC73352.1; -
 DR EMBL: EG14283; YKFF.
 KW Hypothetical protein.
 SO SEQUENCE 79 AA; 9014 MW; F01375F7 CRC32;
 YKFF_ECOLI Length: 79 February 14, 2000 08:03 Type: P Check: 8141

1 MTSOVLPPG PFTROQAV TTYSNITL DQGSFRLV VRDREGMW
 51 RANNEPDAG EGLNRYIRS GIRDTMTR

IIA_SEQUENCE 1.0
 ID YN27_MARPO STANDARD: PRT; 69 AA.
 AC P38469;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last annotation update)
 DE HYPOTHEETICAL 7.9 KD PROTEIN IN NAD6-NAD3 INTERGENIC REGION (ORF 69).
 GN YKFF.
 OS Marchantia polymorpha (Liverwort).
 CC Mitochondrion.
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiopsida;
 CC Marchantiales; Marchantiaceae; Marchantia.
 [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE: 92114051.
 RA ODA K., YAMATO K., OHTA E., NAKAMURA Y., TAKEKURA M., NOZATO N.,
 RA AKASHI K., KANGAE T., OGURA Y., KOHCHI T., OHYAMA K.,
 RT "Gene organization deduced from the complete sequence of liverwort
 RT Marchantia polymorpha mitochondrial DNA. A primitive form of plant
 RT mitochondrial genome."
 RL J. Mol. Biol. 223:1-7(1992).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M68929; AAC09407.1; -
 DR PIR: S25969; S25969.
 DR MENDEL: 2099; MARPO: YN27.1.
 KW Mitochondrion; Hypothetical protein.
 SO SEQUENCE 69 AA; 7940 MW; 699EE385 CRC32;
 YN27_MARPO Length: 69 February 14, 2000 08:03 Type: P Check: 7679

1 MASLSTPRT QTLVPAHTE IYPAHSHF GPAFGQALH LFGPLGLFE
 51 RENACHIFD NRKTPRLQ

IIA_SEQUENCE 1.0
 ID YN06_YEAST STANDARD: PRT; 139 AA.
 AC P53842;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE HYPOTHEETICAL 15.5 KD PROTEIN IN P1K1-POL2 INTERGENIC REGION.
 GN YN126W OR N0800.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
 CC Saccharomycetaceae; Saccharomyces.
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / FY1679;
 RX MEDLINE: 96310631.
 RA SEN-GUPTA M., LYCK R., FLEIG U., NIEDENTHAL R.K., HEGEMANN J.H.,
 RT "The sequence of a 24,152 bp segment from the left arm of chromosome
 RT XIV from Saccharomyces cerevisiae between the BNI1 and the POL2
 RT genes."
 RL Yeast 12:505-514(1996).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X92494; CA63232.1; -
 DR EMBL: Z71542; CA96173.1; -
 KW Hypothetical protein; Transmembrane.
 FT DOMAIN 25 33 POLY-SER.
 FT TRANSMEM 35 55 POTENTIAL.
 FT TRANSMEM 57 77 POTENTIAL.
 SO SEQUENCE 139 AA; 15510 MW; 11A8B5FA CRC32;
 YN06_YEAST Length: 139 February 14, 2000 08:03 Type: P Check: 7226

1 MMLNHTYKL LSYFLRKASN RFEWSSSSF SCSEFLVFLV VFEDCRFSI
 51 TSFLISFGIL SSFLIFSLFC LGFLTVIGCL ASALSLSIS KAKIGFSSSL
 101 SSISPESSLK SEEMLEDD KESSLLYGT STVFAISRK

IIA_SEQUENCE 1.0
 ID YN11_CABEL STANDARD: PRT; 170 AA.
 AC P34496;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE HYPOTHEETICAL 19.1 KD PROTEIN PAR2.1 IN CHROMOSOME III.
 GN K02D10.5 OR PAR2.1.
 OS Caenorhabditis elegans.
 CC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 CC Rhabditina; Rhabditidae; Pelodierinae; Caenorhabditis.
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA WILSON R.;
 RT Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RP SEQUENCE OF 53-170 FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE: 94150718.
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BEKKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPESEY T., COOPER J., COULSON A.,
 RA CAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,


```

251 HALKEVLEP OVALRLADTK AOEKVALNQ FLELMSTEDP RAFYGNHVN
301 RANOELEIET LIAVADSLFRA ODIEITRRKY RVESVREON GKVIFFSMH
351 VSGEOLAQLT GCAILIRFPM PDLDDPMDE N

!!A-SEQUENCE 1.0
ID YOTC_CAEEL STANDARD: PRT: 232 AA.
AC P19296:
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DE HYPOTHETICAL 26.8 KD PROTEIN.
OS Thermoproteus tenax virus 1 (strain KRA1) (TV1).
OC Viruses; dsDNA viruses, no RNA stage; Lipothrixviridae;
OC Lipothrixvirus.
OC [1]
RP SEQUENCE FROM N.A.
RA NEUMANN H.;
RL Submitted (MAR-1989) to the EMBL/Genbank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X14855; CAA32992.1; -
CC DR Hypothetical protein.
CC KW SEQUENCE 232 AA; 26862 MW; E457433C CRC32;
CC
YOTC_TV1 Length: 232 February 14, 2000 08:03 Type: P Check: 9154

1 MDVAVALYS RDDALDLIR TLHYWKRSD GLRGQIYVS NTNVLPRVA
51 KLIOOKRKGV GGARCDALIES TSEDYIIFSD GHVTPPREIS KMLVEPWSV
101 PINHILPFG TSGRAYSLP FMPDEROFLM CSCSKYKEEI TTGGEIVAM
151 SKRIYSMSKC YTSYGLDLFQ FTLRMPGOL IGDEGIHFIE AKRLISNRKP
201 NKDEMKDFYD TIYIKRKEI EEEIKECFCR AL

!!A-SEQUENCE 1.0
ID YOTC_CAEEL STANDARD: PRT: 161 AA.
AC Q10120:
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE HYPOTHETICAL 18.2 KD PROTEIN ZK632.13 IN CHROMOSOME III.
GN ZK632.13.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilia; Rhabdilitia;
OC Rhabdilitia; Rhabdilitidae; Peloderinae; Caenorhabditis.
OC [1]
RP SEQUENCE FROM N.A.
RA STRAIN-BRISTOL N2;
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRATON L., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
RA FULSTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
RA JOHNSON L., JONES M., KERSHAW J., KIRSTEN J., LAISTER N.,
RA LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS E., SHOWNKEEN R.,
RA SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
RA STULTON J., THIERRY-MING J., THOMAS K., VAUGHAN K.,
RA WATKINSON P., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA WOLFDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";

```

```

Nature 368:32-38(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z22181; CAA80188.1; -
CC DR WORMPEP; ZK632.13; CE01711.
CC KW Hypothetical protein.
CC SEQUENCE 161 AA; 18180 MW; 9298BA3 CRC32;
CC
YOTC_CAEEL Length: 161 February 14, 2000 08:03 Type: P Check: 9212

1 MSRLGFTGY EFGDEMEFYQ QMEIKSNAE QAKMEQCK MLECTETMPE
51 ESEVPVAKCL DPEAFQSES VSKCYESPX NISFLKEDAV TVNMSCPA
101 DDIAKLIRNI QNSYTLGIE EARGCRGRL LNVLPFGSA SPRLOFTTP
151 KNAVEETGS Q

!!A-SEQUENCE 1.0
ID YOXD_BACSU STANDARD: PRT: 238 AA.
AC P14803:
DT 01-APR-1990 (Rel. 14, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUN-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL OXIDOREDUCTASE IN RFP-PELB INTERGENIC REGION (EC 1.-.-.-)
DE (ORF238).
OS YOXD.
GN Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OC [1]
RP SEQUENCE FROM N.A.
RA STRAIN-168;
RX MEDLINE: 91192601.
RA ANN K.S., WAKE R.G.;
RT "Variations and coding features of the sequence spanning the
RT replication terminus of Bacillus subtilis 168 and W23 chromosomes.";
RT Gene 98:107-112(1991).
RL [2]
RP SEQUENCE OF 62-238 FROM N.A.
RC STRAIN-168;
RX MEDLINE: 88040469.
RA CARRIGAN C.N., HARRSWA J.A., SMITH M.T., WAKE R.G.;
RT "Sequence features of the replication terminus of the Bacillus
RT subtilis chromosome.";
RL Nucleic Acids Res. 15:8501-8509(1987).
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC FAMILY (SDR).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X06168; CAA29533.1; -
CC DR EMBL: Z99114; CAB13743.1; -
CC PIR: S01270; S01270.
CC HSSP: P19992; 2HS0.
CC SUBTILIST; BG11048; YOXD.
CC PROSITE: PS00061; ADH_SHORT; 1.
CC PFAM: PF00106; adh_short; 1.
CC KW Hypothetical protein; Oxidoreductase.
CC NP_BIND 10 34 NAD OR NADP (BY SIMILARITY).

```

FT ACT_SITE 155 155 BY SIMILARITY
 SQ SEQUENCE 238 AA: 25299 MW: 06381861 CRC32:
 YPOD_BACSU Length: 238 February 14, 2000 08:03 Type: P Check: 2171 ..
 1 MOSLOHKTAL ITGGGGRIGR ATALALAKEG VNIGLIGRTS ANEYKAEV
 51 KALGVKAAPA AADVKADADOV NOAVAOVKEQ IGDIDILINN AGISKEGFL
 101 DLSADEMENI IOVNLGVYH VTRAVLPEMI ERKAGIINI SSTAGORCAA
 151 VTSVYSASKF AVLGTESIM QEVKRNIRY SALTPESTVAS DMSIELULTD
 201 GNEPKVMOPE DLAEYMAQL KIDPRIFIKT AGLMSTNP
 !!AA_SEQUENCE 1.0
 ID YPOD_ECOLI STANDARD: PRT: 285 AA.
 AC 15-DEC-1998 (rel. 37, Created)
 DT 15-DEC-1998 (rel. 37, Last sequence update)
 DE 15-DEC-1998 (rel. 39, Last annotation update)
 GN HYPOTHEICAL TRANSCRIPTIONAL REGULATOR IN DDG-GLK INTERGENIC REGION.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE: 97426617.
 RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
 RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
 RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
 RA MAU B., SHAO Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12.
 RX MEDLINE: 97349980.
 RA YAMAGOTO T., AIBA H., BABA T., HAYASHI K., INADA T., ISONO K.,
 RA ITOH T., KIMURA S., KITAGAWA M., MAKINO K., MITI T., MITSURASHI N.,
 RA MIZOBUCHI K., MORI H., NAKADE S., NAKAMURA Y., NASHIMOTO H.,
 RA OSHIMA T., OYAWA S., SATTO N., SAMPEI G., SATOH Y., SIVASUNDARAM S.,
 RA TAGAMI H., TAKAHASHI H., TAKEEDA J., TAKEKOTO K., UEHARA K., WADA C.,
 RA YAMAGATA S., HORICHI T.;
 RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
 RT -K12 genome corresponding to 50,0-68.8 min on the linkage map and
 RT analysis of its sequence features.";
 RL DNA Res. 4:91-113(1997).
 CC -1- SIMILARITY: BELONGS TO THE ARAC/YLIS FAMILY OF TRANSCRIPTIONAL
 CC REGULATORS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE000326; AAC75441.1; -;
 CC EMBL: D90868; CAB22182.1; -;
 CC DR ECGENE; EGI4150; YPOD.
 CC DR PROSITE; PS00041; HTH_ARAC_FAMILY_1; 1.
 CC DR PROSITE; PS01124; HTH_ARAC_FAMILY_2; 1.
 CC DR PFAM; PF00165; HTH_2; 1.
 CC KW Hypothetical protein; Transcription regulation; DNA-binding.
 CC FT DNA_BIND 200 219 H-T-H motif (BY SIMILARITY).
 CC SQ SEQUENCE 285 AA: 32355 MW: 61FPA84 CRC32:
 YPOD_ECOLI Length: 285 February 14, 2000 08:03 Type: P Check: 4986 ..

1 MKARGLPADQ QRFADLFSGL VLNQLLGRV WFAQSAPASLP VESLCIDFPR
 51 LDIYRGEGY NLEAKOQRL VEGEMLFIPA RANLPVNNK PYMLSLVFA
 101 PTWGLSFYD SRTSLIHPA ROIQLPSPQR GESEAMITAL TILSPLPEQ
 151 NIIOPLVSL LHCRSVYNN PPGNSOPROD ELVHSICNNV QNNYKOPLTR
 201 ESWAGFNIT PNHLSKLFQA HGTMRFIEYV RWMAMAYAM ILOKTHLSIH
 251 EYVQRGEPD SDYFCRVFR OGLTPGEYS ARPOG
 !!AA_SEQUENCE 1.0
 ID YPOD_BACSU STANDARD: PRT: 224 AA.
 AC P42981.
 DT 01-NOV-1995 (rel. 32, Created)
 DT 01-NOV-1995 (rel. 32, Last sequence update)
 DE 15-DEC-1998 (rel. 37, Last annotation update)
 GN HYPOTHEICAL 24.8 KD PROTEIN IN DAPB-PAPS INTERGENIC REGION.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/staphylococcus group; Bacillus.
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168 / MAREBURG;
 RX MEDLINE: 96349105.
 RA SOROKIN A.V., AZEVEDO V., ZUMSTEIN E., GALLERON N., EHRLICH S.D.,
 RA SERROR P.;
 RT "Sequence analysis of the Bacillus subtilis chromosome region between
 RT the sera and kds1 loci cloned in a yeast artificial chromosome.";
 RL Microbiology 142:2005-2016(1996).
 CC -1- SIMILARITY: TO E.COLI YLIS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L38424; AAA92876.1; -;
 CC DR EMBL: L47709; AAB38444.1; -;
 CC DR EMBL: J99115; CAB14163.1; -;
 CC DR SUBTILIST; BG11212; YPOD.
 CC KW Hypothetical protein.
 CC SQ SEQUENCE 224 AA: 24806 MW: DF897440 CRC32:
 YPOD_BACSU Length: 224 February 14, 2000 08:03 Type: P Check: 3316 ..
 1 MYNADVLARG AHSDDVEIGM GGTIAKVRQ EKKYVICDLT EALLSNGTV
 51 SLRKEEAPAA ARILGADKRI QLTLPDRGLI MSDQAIRSIV TVIRICPRPA
 101 VEMPYRKDRH PDHGNAAALV EEAIFSAIGH KYKDEKSLPA HKVSKVYVYV
 151 INGFHPDRV IDISDTIEAK KQSLNAYKSO FIPKDSVST PLTNGYIEIV
 201 EAREKLYGKE AGVEYAEVSF PNCC
 !!AA_SEQUENCE 1.0
 ID YPOD_KLEPN STANDARD: PRT: 271 AA.
 AC P27569.
 DT 01-AUG-1992 (rel. 23, Created)
 DT 01-AUG-1992 (rel. 23, Last sequence update)
 DE 15-DEC-1998 (rel. 37, Last annotation update)
 GN HYPOTHEICAL PROTEIN IN POOA 5' REGION (ORF X) (FRAGMENT).
 OS Klebsiella pneumoniae.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Klebsiella.
 [1]
 RP SEQUENCE FROM N.A.

```

RC STRAIN-NCTC 418:
RC MEDLINE: 92212293.
RX MEULEBERG J.U.M., SELLINK E., RIEGMAN N.H., POSTMA P.W.:
RX Nucleotide sequence and structure of the Klebsiella pneumoniae ppg
RT Operon.
RL Mol. Genet. 232:284-294(1992).
CC -1- FUNCTION: NOT KNOWN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M19.
CC -1- SIMILARITY: TO AN ORF IN THE 3 REGION OF PQO-III IN
CC A.CALCONECTICS.
CC -----
CC THE SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X58778: CAA41578.1: -.
DR PIR: S20452: S20452.
DR PROSITE: PS00869: RENAL_Dipeptidase: 1.
DR Hypothetical protein.
FT NON_TER
FT SEQUENCE 271 AA: 29484 MW: 135841F CRC32:
SQ YPOQ_KLEPN Length: 271 February 14, 2000 08:03 Type: P Check: 9681 ..

1 SVPEDPIDIL WQQLALIKOL IAHSGRLRL CLSADIENC REDKYLAVYA
51 HIEGAGGFDG EGRDLOAFYA AGVNSICPFW NIANRGSGV NGSEFGSDPT
101 GPGTLAAGID LIRQNALAKM QIDVSHNEK AFWDTAHHAT SPLVATHSNA
151 HALCPORNL TDOQLRAIRD SGGVGVNFG NAFRLADGR DSDPLTTIV
201 RHIDYLINIM GEDHVALGSD FDGITLPDEL GDVAGLPRLI NTLRAGYDQ
251 LVLDKLMRN WLRVKNVQ Q

11AA SEQUENCE 1.0
ID YOCE_ECOLI STANDARD: PRT: 425 AA.
AC P77031: P76634; P76635;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHEICAL 46.8 KD PROTEIN IN CYST-ENO INTERGENIC REGION.
GN YOCE
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE: 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.:
RA "The complete genome sequence of Escherichia coli K-12."
RT Science 277:1453-1474(1997).
RL [2]
RN SEQUENCE FROM N.A.
RP STRAIN-K12:
RC ATIA H., BABA T., FUJITA K., HAYASHI K., HONTO A., HORIUCHI T.,
RC IKEMOTO K., INADA T., ISONO K., ISONO S., ITOH T., KANA K.,
RC KASAI H., KASHIMOTO K., KIM S., KIMURA S., KITAGAWA M.,
RC KITAKAWA M., MAKINO K., MASUDA S., MIKI T., MIZOBUCHI K., MORI H.,
RC MOTOMURA K., NAKAMURA F., NASHIMOTO H., NISHIO Y., OSHIMA T.,
RC SAITO N., SAEPEI G., SEKI Y., TAGAMI H., TAKEMOTO K., WADA C.,
RC YAMAMOTO Y., YANO M.:
RC Submitted (Feb-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

```

```

CC -----
CC -1- SIMILARITY: TO E. COLI Y1H.
CC -----
CC THE SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000361: AAC75817.1: -.
DR EMBL: D90893: CAB22504.1: -.
DR ECOGENE: E613174: YOCE.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 9 29
FT TRANSMEM 49 69 POTENTIAL.
FT TRANSMEM 76 96 POTENTIAL.
FT TRANSMEM 97 117 POTENTIAL.
FT TRANSMEM 139 159 POTENTIAL.
FT TRANSMEM 172 192 POTENTIAL.
FT TRANSMEM 220 240 POTENTIAL.
FT TRANSMEM 260 280 POTENTIAL.
FT TRANSMEM 292 312 POTENTIAL.
FT TRANSMEM 314 334 POTENTIAL.
FT TRANSMEM 355 375 POTENTIAL.
FT TRANSMEM 389 409 POTENTIAL.
FT SEQUENCE 425 AA: 46830 MW: 886415D CRC32:
SQ YOCE_ECOLI Length: 425 February 14, 2000 08:03 Type: P Check: 1867 ..

1 MOHNSYRMI TLAIISFSG VSFDLAYLRY IYQPMKFM GFENTIEGLI
51 MSTFGIAIIT LYAPSGYAD KESHRNITS AMITGLGLG LMAITYPLWV
101 MLCIOTAFPI TTILMNSVS IKASLGDH SEQCKIMGM EGRGVGVS
151 LAVFTWVES RPAPDDSTL KVIYIISVY YILGILCWF FVSDNNLRS
201 ANNEKOSFO LSDILAVLRI STWYCSNVI FGVEFTYAIL SYSTNYLTEN
251 YGMSLVASY MGIVINKIFR ALGCPGLGII TTSKVSPT RVLIQLSVLG
301 LRTLTLVYT NSNPQVAMG IGLILGLFT CYASRGLYWA CPGEARTPSY
351 IMGTYGICS VIGFLPDVYV YPIIGHMODT LPAAEVYENM WLMGNALGM
401 VIVFTFLRP KIRVADSAPA MASSK

11AA SEQUENCE 1.0
ID YOCE_ECOLI STANDARD: PRT: 141 AA.
AC P77136:
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHEICAL 16.3 KD PROTEIN IN KDU1-LISS INTERGENIC REGION.
GN YOCE
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE: 97426617.
RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
RA MAU B., SHAO Y.:
RA "The complete genome sequence of Escherichia coli K-12."
RT Science 277:1453-1474(1997).
RL [2]
RN SEQUENCE OF 14-141 FROM N.A.
RP STRAIN-K12 / MG1655;
RC ROBERTS D., ALLEN E., ARAUJO R., APARICIO A., CHUNG E., DAVIS K.,

```

RA DUNCAN M., FEDERSPIEL N., HYMAN R., KALMAN S., KOMP C., KURDI O.,
 LEW H., LIN D., NAWATH A., OEFNER P., SCHAMM S., DAVIS R.W.;
 RA Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF000369; AAC75888.1; ALT_INIT.
 DR EMBL: U83187; AAB40288.1; --
 DR ECGENE: E613274; yjek.
 KM Hypothetical protein.
 SQ SEQUENCE 141 AA; 16271 MW; E96C8D18 CRC32;
 YOEK_ECOLI Length: 141 February 14, 2000 08:03 Type: P Check: 8893 ..
 1 MDIEFSQIHE MYMHDIIVNS DSKKRPRL KFLNAENVL TQTSWTLS
 51 RYVNVSVNK VNKSKYKNS YISRYNDEF SLTDEINSE KETLVLSID
 101 SLSKVLNMP LSVLFTSTVR RNNRKNVY EFDWICTRC C
 11AA_SEQUENCE 1.0
 ID YR7B_ECOLI STANDARD; PRT; 169 AA.
 AC P21316;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 01-MAY-1991 (Rel. 18, Last annotation update)
 DE HYPOTHEICAL 18.1 KD PROTEIN (ORF6) (RETRON EC67).
 OS Escherichia coli.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 CC [1]
 CC SEQUENCE FROM N.A.
 RC STRAIN-CL-1;
 RX MEDLINE; 91067724.
 RA HSU M.-Y., INOUE M., INOUE S.;
 RT "Retron for the 67-base multicopy single-stranded DNA from
 RT Escherichia coli: a potential transposable element encoding both
 RT reverse transcriptase and Dam methylase functions.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9454-9458(1990).
 CC -1- SIMILARITY: 70% IDENTITY TO CP76 OF BACTERIOPHAGE 186.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M55249; AAA23393.1; --
 DR PIR: J00857; J00857.
 SQ SEQUENCE 169 AA; 18129 MW; 9D36865A CRC32;
 YR7B_ECOLI Length: 169 February 14, 2000 08:03 Type: P Check: 6140 ..
 1 MEDYVSKHP HFDACAGFA LKHNVQLAE RAGMNVQLR NKNPQPHL
 51 LTAPIWLLT DLEDESLIV GFLAQHCLP CVPINEVAKE KLPHYMSAT
 101 AEIGVANAAG VSGDVKTSAG RDAIASSINS VTRLMALAV SIQARLOANP
 151 AMAAVDTVT GLGASFGLL
 11AA_SEQUENCE 1.0
 ID YRBC_HAEIN STANDARD; PRT; 214 AA.
 AC P45028;

DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE HYPOTHEICAL PROTEIN H11084 PRECURSOR.
 GN H11084.
 OS Haemophilus influenzae.
 CC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 CC Haemophilus.
 CC [1]
 CC SEQUENCE FROM N.A.
 RC STRAIN-RD / KW20.
 RX MEDLINE; 95350650.
 RA FIEDSCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
 RA KERLAVAGE A.R., BOLT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
 RA MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCKYNE J.D.,
 RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODER A., KELLEY J.M.,
 RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLAD E., COTTON M.D.,
 RA UTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDER D.M., BRANDON R.C.,
 RA FINE L.D., FRITZMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,
 RA GNEHM C.L., McDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
 RA VENTER J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus
 RT influenzae Rd.";
 RT Science 269:496-512(1995).
 CC -1- SIMILARITY: STRONG, TO E.COLI YRBC.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U13788; AAC2740.1; --
 DR TIGR: H11084; --
 KW Hypothetical protein: signal.
 FT SIGNAL
 FT CHARIN 29 214
 SQ SEQUENCE 214 AA; 24510 MW; 02256B8A CRC32;
 YRBC_HAEIN Length: 214 February 14, 2000 08:03 Type: P Check: 1216 ..
 1 MNLQKRMF TLTEVLNLF LVTRTAIET SPVYLMQQA DLFSDIOAN
 51 QSKIKODPNY LRTIVNDLL PYVNLVYGS KVLGSYYKST SAOREKEFF
 101 TFGELIBCKY AQALETYSNQ KIOIESEKEL GDNNFNINRY NIIQANGVAP
 151 ILTFKVRKG NKSGEKVID MVGAGVSMLE DTRKNVGLI NKOIGDTLIT
 201 KMOOSASOPI IFNO
 11AA_SEQUENCE 1.0
 ID YRKG_BACSU STANDARD; PRT; 36 AA.
 AC P54434;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE HYPOTHEICAL 4.2 KD PROTEIN IN BLTR-SPOIIC INTERGENIC REGION.
 GN YRKG
 OS Bacillus subtilis.
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Bacillus.
 CC [1]
 CC SEQUENCE FROM N.A.
 RC STRAIN-168 / JH642;
 RA KOBAYASHI Y., MIZUNO M., MASUDA S., TAKEKAWA K., HOSONO S.,
 RA SATO T., TAKEUCHI M.;
 RA Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
CC EMBL: D84432; BAAL1362.1; -
CC DR EMBL: 299117; CAB14593.1; -
CC DR SUBMITTER: BG1173; YKXG.
CC KW Hypothetical protein.
CC SO SEQUENCE 36 AA; 4234 MW; C3BF5DEA CRC32;

YKXG_BACSU Length: 36 February 14, 2000 08:03 Type: P Check: 1724 ..

1 MYKAMPRE VTKSLTKNL YLFMGCYMKM ILVIYR

!!NA_SEQUENCE 1.0
ID YKXG_BACSU STANDARD; PRT; 118 AA.
AC P85085;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE HYPOTHEICAL 13.8 KD PROTEIN RV3073C.
GN RV3073C OR MTCY22D7.08.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RC MEDLINE: 98295987.
RA COLE S.T., BROSCHE R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA GORDON S.V., EIGLEMEIER K., GAS S., BARRY C.E. III, TERAIA P.,
RA BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
RA DAVIES R., DEVLIN K., FELTWELL T., GENTLES S., HAMILIN N., HOLROYD S.,
RA HORNSBY T., JAGELS K., KROCH A., MCLEAN J., MOLE S., MORPH L.,
RA OLIVER S., OSBORNE J., QUAIL M.A., RAYANDREAN M.A., ROGERS J.,
RA RUTHER S., SEEGER K., SKELTON S., SQUARIS S., SOKES R., STOLSTON J.E.,
RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RT Nature 393:527-544(1998).
RL 1. SIMILARITY: TO E.COLI YKAO.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: 283866; CAB06250.1; -
CC KW Hypothetical protein.
CC SO SEQUENCE 118 AA; 13750 MW; B38D283F CRC32;

YU73_MYCTU Length: 118 February 14, 2000 08:03 Type: P Check: 3649 ..

1 MYRTRVRA RYEDIDPD GORVLDRIH PHGIRKQOR VGIMCKRYAP

51 SKLEWYHH OPRFDEFAS RQDELHDSA ALAEIRKLTG RSVYTPVAT
101 RHVAASHAAV LAOLLNR

!!NA_SEQUENCE 1.0
ID YU73_MYCTU STANDARD; PRT; 394 AA.
AC P36442;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE HYPOTHEICAL 43.4 KD PROTEIN IN ROCC-PTA INTERGENIC REGION.
KW YU73 OR IPA-84D.
RC Bacillus subtilis.

CC Bacteria; Firmicutes; Bacillus/Clostridium group;
CC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RC MEDLINE: 95020537.
RA GLASER P., KUNST F., ARNOLD M., COUDART M.P., GONZALES W.,
RA HULLO M.F., IONESCUCU M., LUBCHINSKY B., MARCELINO L., MOSZER I.,
RA PRESECAN E., SANTANA M., SCHNEIDER E., SCHWEIZER J., VERTES A.,
RA RAPOPORT G., DANCHIN A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
RT Mb region from 325 degrees to 333 degrees."
RL Mol. Microbiol. 10:371-384(1993).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: X73124; CA51640.1; -
CC DR EMBL: 298123; CAB15797.1; -
CC DR PIR: S39739; S39738.
CC SUBMITTER: BG10630; YWFF.
CC KW Hypothetical protein; Transmembrane.
CC FT TRANSMEM 11
CC FT TRANSMEM 31
CC FT TRANSMEM 43
CC FT TRANSMEM 63
CC FT TRANSMEM 69
CC FT TRANSMEM 92
CC FT TRANSMEM 142
CC FT TRANSMEM 166
CC FT TRANSMEM 215
CC FT TRANSMEM 244
CC FT TRANSMEM 288
CC FT TRANSMEM 332
CC FT TRANSMEM 353
CC FT TRANSMEM 373
CC SO SEQUENCE 394 AA; 43416 MW; 93D522A2 CRC32;

YWFF_BACSU Length: 394 February 14, 2000 08:03 Type: P Check: 3644 ..

1 MKOLKPNKXK LLYGALSFM GDYCVLPALL ILSTYHDW VTSQVIYRS

51 IPVFPQPIG VLVDRDLRIK IMLTDLIRG IIFLGITPLP KGEYPLFLA
101 LFTITGSGV FENPARLAVM SLESDIKSI NTLPAKATTI STIYGAAAG
151 LFLIGSVEL AVAFNGVYLL VSAFISRIK LQFVIOSEN IKAEPQSFKE
201 GLKEIKTNSF VLNAMFTMIT MALMGVYVS YPIYSPRLG DGEIGNFILT
251 FCGFGGFTG AALYSKMGFN NNRGLYFTV LSIIVLAFU FTPIFAVSI
301 AALIFIANE YGEVLAKYV QENANQIOG RIFSVAESI GLCISISGMF
351 INLSAPVIM GLIIVYGL FLHTKLVRK FLERDNKTEQ KGYV

!!NA_SEQUENCE 1.0
ID YWFF_BACSU STANDARD; PRT; 274 AA.
AC Q93834;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHEICAL 31.6 KD PROTEIN F596.8 IN CHROMOSOME I.
GN F596.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Secernentea; Rhabdita; Rhabditidae;
OC Rhabditiina; Rhabditidae; Peloderiinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL NZ;

RA WILKINSON J.:
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO A FAMILY OF C.ELEGANS PROTEINS THAT GROUPS
 CC C37A7.2, C33H5.2, C35A5.5, F13G3.3, F59C6.8, FA9C12.5, R07B7.12
 CC AND ZK381.2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z79600; CAB01879.1; -
 DR WORMPEP: F59C6.8; CE11470.
 KW Hypothetical protein.
 SQ SEQUENCE 274 AA; 31624 MW; 43119A02 CRC32;
 YYP_CAEEL Length: 274 February 14, 2000 08:03 Type: P Check: 1176 ..
 1 MGNVEVOFK IOPVINTIPE FCKWVPYLAV GQVEDHVLV KLSTNKIDGM
 51 ELSRTPEYET PRKVACFSP LFUNERWOLL LATVEIYSHY GAFMHEYTRI
 101 SPMSAIRIGE SRAASPMFDP NTELEFRNOA SAMTDLLOX KEAEFIVFP
 151 DPDDLVPVL GKNYEEFTQ AFKMEPTAGA VYVNTQTSI ESSMTPALYS
 201 PISMLASMKF KGEQKMGKLV VPERVDSTW IHRSAIKES FQKVMFVDV
 251 NAFYLRIRWK FPEVPTFNRS SELL
 11AA_SEQUENCE 1.0 STANDARD; PRT: 364 AA.
 ID ZRP4.MAIZE
 AC P47917;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE O-METHYLTRANSFERASE ZRP4 (EC 2.1.1.-) (OMT).
 GN ZRP4.
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
 OC Poaceae; Zea.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV, NKH31; TISSUE=ROOT;
 RX MEDLINE: 94105316.
 RA HEID B.M., WANG H., JOHN I., WURTELE E.S., COLBERT J.T.;
 RT "An mRNA putatively coding for an O-methyltransferase accumulates
 RT preferentially in maize roots and is located predominantly in the
 RT region of the endodermis."
 RL Plant Physiol. 102:1001-1008(1993).
 CC -1- FUNCTION: MAY BE INVOLVED IN THE O-METHYLATION OF SUBERIN
 CC PHENYLPROPANOID PRECURSORS.
 CC -1- TISSUE SPECIFICITY: ACCUMULATES PREFERENTIALLY IN THE ROOTS AND IS
 CC LOCATED PREDOMINANTLY IN THE REGION OF THE ENDODERMIS, LOW LEVELS
 CC ARE SEEN IN THE LEAVES, STEMS, AND OTHER SHOOT ORGANS.
 CC -1- SIMILARITY: TO OTHER OMTS REQUIRING S-ADENOSYL-L-METHIONINE AS
 CC SUBSTRATE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L14063; AAA18532.1; -
 DR MAIZEDB: 63528; -
 DR PFM: PFM0891; Methyltransf_2; 1.

KW Transferase; Methyltransferase.
 SQ SEQUENCE 364 AA; 39583 MW; 3067DB07 CRC32;
 ZRP4_MAIZE Length: 364 February 14, 2000 08:03 Type: P Check: 6645 ..
 1 MEISPNNSRD QSLDAQLEL WHTTFAPMKS MALKSAHLIR IADAIHNGG
 51 AASLSQILSK VLNHSRYSS LRRLMVLVT TVEFGTPLG GGSDDSEFP
 101 YTLTPYSRLI IGSOSQIAQ TPLAAVLDP TIVSPFSELG AMFOHELPPD
 151 CTRKHTGRG IWELTKDPT FDLAVNDGIA SDSQILYDVA IKOSAEVFOG
 201 ISSLVVGGG IGAAQAISK APFVKCSVL DIAHVAKAP THTDVOPIAG
 251 DMFESIPPAD AVLKSVLHD WDHDCVKIL KNCKRAIPPR EAQGVYIIN
 301 MVMGAPSDM KHKWQALFD VYIMFNGME RDEDEWSKIF SEAGYSYRI
 351 IPVLGVRSII EYYP

